



# Generación de resúmenes audiovisuales a partir de obras literarias utilizando análisis de emociones

Milón Flores Daniela Fernanda

Orientador: Dr. Erick Gomez Nieto

*Tesis profesional presentada a la Escuela Profesional de Ciencia de la Computación como parte de los requisitos para obtener el Título Profesional de Lic. en Ciencia de la Computación.*

UCSP- Universidad Católica San Pablo  
Agosto de 2019

*Dedico esta tesis a mis padres, quienes a lo largo de los años, me han guiado con su consejo y han protegido mi corazón. Con este trabajo, un capítulo de mi vida se cierra, pero una nueva aventura comienza, y sé que en esta también los tendré junto a mí.*

# Abreviaturas

**AAC** *Affective Algorithmic Composition*

**AI** *Artificial intelligence*

**AC** *Algorithmic Composition*

**SBR** *Sistemas-Basados en Reglas*

**RNA** *Redes Neuronales Artificiales*

# Resumen

---

La lectura de obras literarias es una actividad esencial para la comunicación y el aprendizaje humano. Sin embargo, varias tareas relevantes como la selección, el filtrado o el análisis en un gran número de obras se vuelven complejas. Para hacer frente a este requisito, se proponen varias estrategias para inspeccionar rápidamente cantidades sustanciales de texto, o recuperar información previamente leída, explotando los datos gráficos, textuales o auditivos. En este trabajo, proponemos una metodología para generar resúmenes audiovisuales mediante la combinación de una composición musical basada en emociones y una animación basada en grafos. Aplicamos algoritmos de procesamiento de lenguaje natural para extraer emociones y personajes involucrados en la obra literaria. Luego, utilizamos la información extraída para componer una pieza musical que acompaña la narración visual de la historia con el objetivo de transmitir la emoción extraída. Para ello, fijamos características musicales importantes como progresión de acordes, tempo, escala y octavas, y asignamos un conjunto de instrumentos que se adapte mejor a cada emoción. Además, animamos un grafo para resumir los diálogos entre los personajes de la obra. Finalmente, para evaluar la calidad de nuestra metodología, realizamos dos estudios con usuarios que revelan que nuestra propuesta proporciona un alto nivel de comprensión sobre el contenido de la obra literaria además de aportar una experiencia agradable al usuario.



# Abstract

---

Literature work reading is an essential activity for human communication and learning. However, several relevant tasks as selection, filter or analyze in a high number of such works become complex. For dealing with this requirement, several strategies are proposed to rapidly inspect substantial amounts of text, or retrieve information previously read, exploiting graphical, textual or auditory resources. In this work, we propose a methodology to generate audio-visual summaries by the combination of emotion-based music composition and graph-based animation. We applied natural language processing algorithms for extracting emotions and characters involved in literary work. Then, we use the extracted information to compose a musical piece to accompany the visual narration of the story aiming to convey the extracted emotion. For that, we set important musical features as chord progressions, tempo, scale, and octaves, and we assign a set of suitable instruments. Moreover, we animate a graph to sum up the dialogues between the characters in the literary work. Finally, to assess the quality of our methodology, we conducted two user studies that reveal that our proposal provides a high level of understanding over the content of the literary work besides bringing a pleasant experience to the user.

# Índice general

<b>1. Introducción</b>	<b>2</b>
1.1. Motivación y contexto . . . . .	3
1.2. Planteamiento del problema . . . . .	3
1.3. Objetivos . . . . .	4
1.3.1. Objetivos específicos . . . . .	4
1.4. Organización de la tesis . . . . .	5
<b>2. Fundamentos teóricos</b>	<b>6</b>
2.1. Consideraciones iniciales . . . . .	6
2.2. Teoría musical . . . . .	6
2.2.1. Nota musical . . . . .	6
2.2.2. Octavas . . . . .	7
2.2.3. Pentagrama musical . . . . .	7
2.2.4. Melodía . . . . .	8
2.2.5. Armonía . . . . .	9
2.2.6. Tempo . . . . .	9
2.2.7. Escalas . . . . .	9
2.2.8. Acordes . . . . .	10
2.2.9. Consonancia . . . . .	10
2.3. Relación entre estructura musical y emociones . . . . .	10
2.3.1. Tempo . . . . .	11

---

2.3.2.	Escalas . . . . .	11
2.3.3.	Volumen o intensidad . . . . .	11
2.3.4.	Octavas . . . . .	11
2.3.5.	Armonías . . . . .	11
2.3.6.	Acordes . . . . .	12
2.3.7.	<i>Pitch</i> o tono . . . . .	12
2.4.	Consideraciones finales . . . . .	12
<b>3.</b>	<b>Trabajos relacionados</b>	<b>13</b>
3.1.	Análisis de emociones en un texto . . . . .	13
3.2.	Composición algorítmica . . . . .	14
3.2.1.	Enfoque basado en gramáticas . . . . .	14
3.2.2.	Enfoque basado en reglas . . . . .	15
3.2.3.	Enfoque basado en las cadenas de Márkov . . . . .	15
3.2.4.	Enfoque basado en redes neuronales artificiales . . . . .	16
3.2.5.	Enfoque evolutivo . . . . .	16
3.2.6.	Algoritmos afectivos para composición . . . . .	17
3.3.	Generación de resúmenes visuales a partir de un texto . . . . .	17
<b>4.</b>	<b>Propuesta</b>	<b>19</b>
4.1.	Extracción de emociones y personajes de obras literarias . . . . .	19
4.2.	Composición musical basada en emociones . . . . .	23
4.2.1.	Módulo de la armonía . . . . .	23
4.2.2.	Módulo del ritmo . . . . .	25
4.2.3.	Módulo de la melodía . . . . .	27
4.3.	Animación del grafo . . . . .	29
<b>5.</b>	<b>Resultados</b>	<b>31</b>

---

---

5.1. Conteo de emociones usando los recursos de: Texto etiquetado y Lexicones emocionales . . . . .	31
5.2. Extracción de características, grafos y pentagramas . . . . .	33
5.3. Estudio con usuarios . . . . .	35
5.3.1. Probando la composición musical . . . . .	35
5.3.2. Preguntas sobre la experiencia del usuario usando nuestro prototipo	35
<b>6. Conclusiones y trabajos futuros</b>	<b>40</b>
6.1. Conclusiones . . . . .	40
6.2. Trabajos futuros . . . . .	41
<b>Bibliografía</b>	<b>46</b>

## Índice de tablas

3.1. Matriz de Márkov para la selección de acordes. Extraído de [Morreale et al., 2013] . . . . .	16
4.1. Emociones básicas y avanzadas. Basado en Plutchik [1984]. . . . .	23
4.2. Lista de instrumentos de la API JFugue asociados a cada emoción . . . . .	29
5.1. Emociones extraídas, actividad, tempo, octavas por sección en cada obra literaria procesada. . . . .	34
5.2. Preguntas sobre la experiencia de usuario en nuestro segundo estudio de usuarios. . . . .	36

# Índice de figuras

2.1. Pentagrama musical con un compás de 4/4. . . . .	8
4.1. <i>Pipeline</i> empleado para nuestra metodología. . . . .	19
4.2. Acordes más populares para escalas mayores y menores según [Findeisen, 2015] . . . . .	24
4.3. Progresiones de acordes seleccionadas para la obra “ <i>Peter Pan and Wendy</i> ”. . . . .	25
4.4. Una visión general de nuestro prototipo: (izquierda) una versión resaltada de los diálogos de texto y (derecha) un grafo animado que ilustra la interacción entre personajes. . . . .	30
5.1. Matriz de confusión que evalúa el conjunto de datos homogéneo. . . . .	32
5.2. Extracción de emociones usando dos recursos diferentes: Texto etiquetado (izquierda) y Lexicones emocionales (derecha). Aplicado a las siguientes obras: “ <i>Peter Pan and Wendy</i> ”, “ <i>Psycho</i> ” y “ <i>The Lovely Bones</i> ” respectivamente. Los colores representan las emociones reconocidas: ■ = alegría, ■ = tristeza, ■ = ira, ■ = miedo, ■ = anticipación, y ■ = sorpresa. . . . .	32
5.3. Mostramos los grafos resultantes de haber procesado las cuatro secciones de “ <i>Peter Pan and Wendy</i> ”. . . . .	37
5.4. Representación visual del procesamiento completo de “ <i>Psycho</i> ” por Robert Bloch. . . . .	38
5.5. Resultados obtenidos por nuestro primer estudio de usuarios acerca de cuatro composiciones musicales diferentes: (a) “ <i>Peter Pan and Wendy</i> ” (alegría), (b) “ <i>Psycho</i> ” (miedo), (c) “ <i>Don Quixote of La Mancha</i> ” (alegría), y (d) “ <i>The three musketeers</i> ” (tristeza). Los colores representan la emoción elegida: ■ = miedo, ■ = ira, ■ = tristeza, y ■ = alegría por cada uno de los 30 participantes. . . . .	39
5.6. Resultados obtenidos por nuestro segundo estudio de usuarios en la primera sección de “ <i>Peter Pan and Wendy</i> ”. Los colores representan el valor elegido: ■ (1), ■ (2), ■ (3), ■ (4), y ■ (5) por cada uno de los 7 participantes. . . . .	39

# Capítulo 1

## Introducción

Hoy en día, la Información o Era Digital en la que vivimos nos proporciona un amplio conjunto de recursos multimedia que nos sirven de apoyo para llevar a cabo tareas rutinarias, como por ejemplo, videollamadas, búsqueda de información, compartir fotos, pedir comida en línea para la entrega, etc. Sin embargo, en el ámbito de la lectura, el desarrollo de recursos que permiten la realización de varias tareas relevantes como la selección, el filtrado o el análisis en un gran número de obras se ha vuelto complejo. Para hacer frente a este requisito, se proponen varias estrategias para inspeccionar rápidamente cantidades sustanciales de texto, o recuperar información previamente leída, explotando los datos gráficos, textuales o auditivos como en los trabajos de [Davis and Mohammad, 2014, Hiller Jr and Isaacson, 1957, Yamada and Murai, 2009]. En particular, algunas de estas estrategias han sido dirigidas a transferir emociones del texto de los libros a la música y a las metáforas visuales.

Así, por un lado encontramos el campo de *Affective Algorithmic Composition* (AAC), que combina la composición musical asistida por ordenador y su evaluación emocional [Williams et al., 2015]. Este campo aborda trabajos que conectan la detección automática de emociones con la generación automática de música. Por otro lado, hubo varios intentos de entender y analizar visualmente el texto de los libros. Las redes topológicas, gráficos, cuadros, líneas de tiempo, por nombrar algunos, se han utilizado para caracterizar y resumir mejor el texto en novelas y poemas [Elson et al., 2010, Mohammad, 2011, Yamada et al., 2013]. Sin embargo, basados en nuestra investigación, sabemos que no hay trabajos que automáticamente compongan una pieza musical basada en emociones y que al mismo tiempo, proporcionen una percepción visual del texto. El objetivo de este trabajo es llenar este vacío y ofrecer una mejor experiencia en la lectura de libros, principalmente novelas. En este sentido, nuestra contribución es doble. En primer lugar, proponemos una metodología para generar resúmenes audiovisuales mediante la combinación de una composición musical basada en emociones y una animación de grafos, que al transmitirse de forma simultánea, se encargan de narrar el contenido de la obra. En segundo lugar, proporcionamos un prototipo de nuestra implementación que nos permite evaluar los resúmenes audiovisuales creados.

Con el fin de generar resúmenes audiovisuales, primero aplicamos algoritmos de procesamiento de lenguaje natural para extraer emociones y personajes involucrados en la obra literaria. Luego, guiados por esas emociones, componemos una pieza musical para acompañar la narración visual de la historia. Con respecto a la composición algorítmica

ca afectiva, se han considerado características musicales importantes como progresión de acordes, tempo, escala y octavas. Además, se ha seleccionado un conjunto de instrumentos adecuados para transmitir las emociones extraídas.

Para crear un resumen visual, desarrollamos una animación basada en grafos para representar cómo interactúan los personajes entre sí. De hecho, contamos y mostramos las frecuencias de variación de cada personaje a lo largo de la historia utilizando los nodos del grafo. Las relaciones entre los personajes están representadas por las aristas, el ancho de estas, denota la frecuencia de interacción entre dos personajes específicos.

De la implementación de nuestro prototipo, nos interesó evaluar su capacidad para resumir y comunicar de manera perceptiva. Nuestros experimentos con varias novelas literarias revelan que nuestra propuesta proporciona un alto nivel de comprensión sobre el contenido de una sección específica o de un libro entero. Además, aporta una experiencia agradable al usuario.

## 1.1. Motivación y contexto

En la actualidad, existe una gran variedad de herramientas multimedia que nos sirven para la realización de actividades cotidianas (consultas virtuales, videollamadas, búsqueda de información). Para muchas de estas herramientas, se necesita de una gran cantidad de datos para cumplir su función. En el ámbito de la lectura, también se enfrentan este tipo de dificultades, sobre todo en tareas como: selección, filtrado o análisis en un gran número de obras literarias. Por otro lado, el amplio conjunto de recursos multimedia que tenemos a nuestra disposición (teléfonos móviles, computadoras, videojuegos), actúan como distractores, impidiendo que el público en general disfrute del contenido de una obra literaria. De esta forma, proponemos desarrollar un prototipo que, a través de los recursos multimedia, nos permita aumentar el compromiso de los usuarios para descubrir en detalle la historia que se esconde detrás de nuestro resumen audio-visual y al mismo tiempo, desarrollar estrategias para inspeccionar rápidamente cantidades sustanciales de texto, y recuperar información previamente leída, explotando los datos gráficos, textuales y auditivos. Finalmente, una vez que se hayan alcanzado todos los objetivos, las aplicaciones que se pueden realizar con nuestro modelo poseen una gran diversidad. Inicialmente podríamos generar resúmenes audiovisuales para cualquier novela literaria y más adelante podría convertirse en un recurso multimedia que incorpore algunas herramientas interactivas para la exploración del usuario y así aumentar su capacidad en el análisis de libros. Y luego sería posible generar resúmenes audio-visuales que no estén ligados meramente a una obra literaria, sino a un libro de historia, al *script* de una película o a los versos de un poema.

## 1.2. Planteamiento del problema

Sabemos por nuestra investigación, que no hay ningún trabajo que componga automáticamente música basada en emociones y que al mismo tiempo proporcione una perspectiva visual a partir de un texto, por lo que el primer problema con el que nos



enfrentamos fue el encontrar información que nos sirva de punto de partida. Además, existen muchos otros factores que generan dificultades, por ejemplo: con respecto a la composición musical basada en emociones, no sólo hay una gran escasez de algoritmos enfocados en la composición afectiva, sino que también hay una gran ausencia de datos asociados a una emoción específica. Por otro lado, los métodos de aprendizaje automático no han alcanzado una alta precisión en sus predicciones, precisamente por la falta de datos. Además de ello, es necesario tener un amplio conocimiento en teoría musical que nos permita establecer reglas para una composición que sea coherente y consonante. Y puesto que, nuestra intención es generar una pieza musical a partir de una obra literaria, es necesaria la realización de un análisis de emociones y la extracción de personajes y diálogos. Actualmente no contamos con muchas herramientas que se enfoquen en la extracción de emociones o en el reconocimiento de entidades en un texto. Aquellos *frameworks* que lo hacen son muy limitados y generan muchos errores, por ejemplo: muchos personajes de los cuentos infantiles no son reconocidos por el *framework* (Ejemplo: Pinocchio o Aladin), se necesita de mucha memoria para ejecutar la herramienta y el tiempo de ejecución, muchas veces es muy tardío. Adicionalmente, es necesario contar con un vasto conjunto de datos de texto (cuentos, guiones, novelas, etc.) para la obtención de un análisis estadístico con fundamento. Otro factor en contra es la ausencia de comparaciones con otros algoritmos en la literatura, pues según lo analizado, cada algoritmo posee parámetros que son aleatorios, ambiguos o vagamente definidos, lo que impide una comparación exacta entre los métodos a estudiar. Finalmente, este no es un modelo que pueda ser validado con facilidad, no existe un método de validación objetivo que se centre específicamente en este tipo de modelos.

## 1.3. Objetivos

Desarrollar una metodología para generar resúmenes audiovisuales mediante la combinación de una composición musical basada en emociones y una animación basada en grafos.

### 1.3.1. Objetivos específicos

1. Implementar un algoritmo para realizar la extracción de emociones a partir de un texto literario.
2. Implementar un método que componga una pieza musical con múltiples instrumentos y de forma automática en base a las emociones obtenidas de la obra literaria.
3. Implementar un método para mapear los personajes y diálogos obtenidos de la obra literaria en un grafo animado que represente la interacción entre los personajes principales de la obra.
4. Implementar un prototipo que combine la animación del grafo con la melodía musical para generar resúmenes audiovisuales.
5. Realizar pruebas del prototipo con distintas obras literarias y ejercer estudios con usuarios para la obtención de *feedback* sobre el prototipo.

---

## **1.4. Organización de la tesis**

Este trabajo está organizado de la siguiente manera. En el Capítulo 2 presentamos los fundamentos teóricos. Los trabajos relacionados se presentan en el Capítulo 3. En el Capítulo 4 se introduce nuestra propuesta para generar resúmenes audiovisuales. El Capítulo 5 está dedicado a los Resultados. Finalmente, el Capítulo 6 concluye el documento.

## Capítulo 2

# Fundamentos teóricos

### 2.1. Consideraciones iniciales

Este capítulo no asume un conocimiento previo en la terminología de teoría musical, por ello hacemos uso de las siguientes secciones para ahondar un poco más en los conceptos clave que nos permitirán entender los siguientes capítulos. Empezaremos describiendo una gran cantidad de elementos de la estructura musical que debemos conocer si deseamos componer una pieza de música. También hablaremos de cómo es que la mayoría de estos elementos tiene una estrecha relación a la hora de inducir emociones en el oyente y utilizaremos esta información a nuestro favor para realizar el algoritmo de composición musical que nuestro modelo emplea.

### 2.2. Teoría musical

La teoría musical es un campo de estudio que tiene por objeto la investigación de los diversos elementos de la música [Kennedy et al., 2006]. Como resultado se han generado diversos conceptos para cada uno de sus elementos y a continuación veremos los más relevantes para este trabajo. Si se desea profundizar en teoría musical Baxter and Baxter [2007], Zamacois [1978] son autores de libros con mucho contenido base que pueden ser de ayuda.

#### 2.2.1. Nota musical

Una nota musical, no es simplemente el nombre de la nota que tocamos en un instrumento. (DO, RE, MI, FA, SOL, LA, SI), sino que ésta se ve compuesta por un tono, una duración, la intensidad y el timbre. Y a estas cuatro se las conoce como las cualidades del sonido [Herrera, 1984].

1. El tono o *pitch* por su nombre en inglés, es la cualidad del sonido por la que reco-

nocemos si éste es agudo o grave.

2. La duración de una nota, se ve ligada al tiempo, son los segundos que dura la vibración sonora. Las duraciones existentes de una nota son: La nota blanca que dura 4 pulsaciones, una nota media que dura 2 pulsaciones, una nota negra que dura 1 pulsación, una nota corchea que dura media pulsación y una nota semicorchea que dura un cuarto de pulsación. También existen las fusas y semifusas. Como vemos, hemos descrito las notas desde las que duran más tiempo hasta las que duran menos tiempo así estas últimas se tocan con más velocidad y se escuchan más rápido, por lo que musicalmente la duración está asociada al ritmo [Herrera, 1984].
3. Luego encontramos a la intensidad o volumen, esta cualidad determina si los sonidos son fuertes o suaves, por ejemplo, podríamos tocar una melodía en guitarra de forma suave, intentando expresar una emoción de calma y pausa, o podemos decidir tocar la misma melodía de una forma más intensa, intentando expresar una emoción con más energía.
4. Finalmente tenemos al timbre, esta es la cualidad que hace que sea posible reconocer a los instrumentos, un ejemplo de ello sería escuchar una canción y percibir que el piano se ve acompañado por una melodía de violín, o escuchar a una primera guitarra y distinguir a una segunda que la acompaña de forma más suave, estos sonidos reciben el nombre de armónicos

### 2.2.2. Octavas

En occidente utilizamos doce sonidos [Zamacois, 1978]. Hay siete sonidos naturales y cinco alterados.

- Notas naturales: DO, RE, MI, FA, SOL, LA, SI
- Notas alteradas: DO#/Reb, RE#/Mib, FA#/SOLb, SOL#/Lab, LA#/Sib

Las Notas alteradas reciben dos nombres. Por ejemplo, DO# (DO mayor) porque es seguida de la nota DO o REb(RE bemol) porque precede a la nota RE. Cuando tocamos las primeras doce notas podemos decir que hemos tocado la primera octava, si tocamos las siguientes doce, habremos tocado la segunda octava y así sucesivamente a lo largo del registro de cada instrumento musical. Es necesario aclarar que las primeras octavas conforman sonidos más graves, mientras que las últimas suenan de forma más aguda.

### 2.2.3. Pentagrama musical

La música como cualquier otro lenguaje, tiene reglas al momento de escribirla y de leerla, y para ello hacemos uso del pentagrama musical.



Figura 2.1: Pentagrama musical con un compás de 4/4.

En la Figura 2.1. podemos observar un Pentagrama musical y los distintos elementos que lo conforman. El pentagrama musical tiene cinco líneas y cuatro espacios. Y a cada nota le corresponde un espacio en el pentagrama. Por ejemplo, la primera nota que observamos en el pentagrama de la Figura 2.1 es la nota MI y es una nota blanca, lo que significa que dura 2 pulsaciones.

A continuación, describiremos brevemente los elementos del pentagrama musical.

1. **Measure:** El pentagrama, se encuentra dividido por *bar lines*, rayas verticales que dividen al pentagrama en *measures* o *bars*, y cada uno de estos contiene una cierta cantidad de notas. ¿Cómo saber cuantas notas entran en un *measure*?, para ello usamos nuestra signatura de compás.
2. **Signatura de compás:** En la Figura 2.1 notamos un número 4 sobre otro número 4, a esta proporción se le llama signatura de compás o *time signature* en inglés. El numerador (el 4 de arriba), nos indica qué:  
El número de notas que pongamos en el *measure* debe dar como resultado una duración de 4 pulsaciones. Por ejemplo:  
Podemos elegir poner en el primer *measure* 4 notas negras (una pulsación cada una), o una nota media ( 2 pulsaciones), 4 semicorcheas (1 pulsación) y una negra (1 pulsación). En ambas propuestas, el resultado nos da una duración total de 4, pero en la primera combinación hay 4 notas y en la segunda hay 6 notas. La signatura de compás de 4/4 es la más común y como vimos, está formada por cuatro pulsos en total (fuerte, débil, débil, débil).

## 2.2.4. Melodía

La melodía es un conjunto de notas y es uno de los elementos más básicos de la música, entonces, si una nota es un sonido con un tono y duración particular, la inserción de una serie de notas juntas, una después de la otra, se le conoce como melodía [Schmidt-Jones and Jones, 2011]. La melodía es la forma de combinar los sonidos, pero sucesivamente. De ahí que a muchos instrumentos se los llame melódicos, por ejemplo, una flauta, un saxo, un clarinete o cualquier instrumento de viento, porque ellos no pueden hacer sonar más de una nota a la vez [Zamacois, 1978]. De igual forma con el tema *Estrellita dónde estás*, si se escucha esta pieza musical, se podrá percibir una serie de notas que se tocan de forma secuencial y que en ningún momento se oyen dos notas sonar al mismo tiempo.

### 2.2.5. Armonía

Usando melodías solamente, los temas sonarían vacíos. Por eso se necesita de algo que adorne la canción, y la haga sentir menos escueta. La armonía es la forma de combinar sonidos en forma simultánea [Zamacois, 1978]. Cada compositor la usará para crear diferentes climas haciéndonos sentir una variedad de sensaciones. Los instrumentos llamados armónicos, como el piano o la guitarra, son los que pueden tocar más de una nota a la vez. Además, las armonías no se limitan a tocar varias notas de forma simultánea de un único instrumento, sino que pueden estar conformadas por varios instrumentos, haciendo que la canción se sienta aún más viva.

### 2.2.6. Tempo

Es la velocidad con la que se toca la pieza, es un término en italiano y significa *tiempo* [Herrera, 1984]. Existen diferentes tipos de tempo, a continuación los más comunes.

- *Adagio* : Lento, 66-76 bpm
- *Moderato* : Moderado, 108-120 bpm
- *Allegro* : Alegre, 120-168 bpm
- *Presto* : Muy rápido, 168-200 bpm
- *Prestissimo* : Rapidísimo, 200-208

Para lograr el *tempo* correcto, se hace uso de un metrónomo que marca los pulso por minuto de manera exacta.

### 2.2.7. Escalas

Ya sabíamos que una octava está compuesta de 12 notas, para formar una escala seleccionamos algunas de ellas [Rowe, 2001]. Por ejemplo la escala de *C major* (DO mayor), se formó seleccionando solo las 7 notas blancas de una octava y curiosamente el piano funciona en base a la escala de DO mayor.

- Escala de DO mayor: DO,RE,MI,FA,SOL,LA,SI
- Escala de DO menor: DO,RE,Mib,FA,SOL,LAb,Sib

Existen otro tipos de escalas como LA Menor, pero estas son las más conocidas.

### 2.2.8. Acordes

Demos recordar que la armonía es el estudio de los acordes y estos son los pilares sobre los que se sustenta la melodía en palabras de Jaime Altozano<sup>1</sup>. Así, los acordes musicales se usan para embellecer y darle forma a una canción, para darle armonía. Tocar tres o más notas al mismo tiempo recibe el nombre de acorde [Rowe, 2001], y existen de varios tipos. Acordes mayores, menores, aumentados, disminuidos, etc. La forma más común de construir un acorde es eligiendo una escala. Suponiendo que escogemos DO mayor. Esta escala tiene 7 notas base, y un acorde se construye tomando la nota base y le sumamos su tercera y quinta. Como tenemos 7 notas base tendremos una serie de 7 acordes (sin séptima). Un ejemplo es tomar la nota base DO, ahora le añadimos MI y SOL, y ese es nuestro primer acorde.

### 2.2.9. Consonancia

Este es el último término clave que analizaremos. La consonancia es la cualidad de aquellos sonidos que cuando se tocan al mismo tiempo, producen un efecto agradable Latham [2009]. Emplearemos esta cualidad más adelante para producir una serie de sensaciones musicales.

Ciertamente no hemos profundizado en todos los conceptos que abarca la Teoría musical, pero estos son suficientes para entender la propuesta de este trabajo.

## 2.3. Relación entre estructura musical y emociones

En esta sección exploramos los diferentes resultados de pruebas que se llevan realizando durante varias décadas y cuyo objetivo es conocer cómo los distintos factores de la estructura musical tienen un impacto en la percepción de emociones del oyente. ¿A qué nos referimos con esto?, muchas veces cuando escuchamos una melodía, podemos percibir casi de inmediato si esta nos produce una sensación de felicidad o de tristeza. De hecho, como bien intenta explicar la filósofa Susanne Langer, cuando un compositor desea escribir música, él sabe lo que desea expresar, conoce la forma de las emociones y puede manejarlas, para finalmente componerlas. Por esta razón, vale la pena revisar cómo es que estos factores permiten que el oyente se conecte con sus emociones con cada pieza musical.

A continuación presentaremos un listado de aquellos factores que han demostrado poseer una estrecha influencia en el aspecto emocional gracias al exhaustivo estudio de autores como Gabrielsson and Lindström [2001], Juslin and Västfjäll [2008], Lindström and Västfjäll [2009].

---

<sup>1</sup>Compositor y productor musical español:<https://jaimealtozano.com/>

### 2.3.1. Tempo

El *tempo*, de entre todos los factores musicales que afectan las expresiones emocionales, ha sido el más estudiado. Investigaciones como las de [Gabrielsson and Lindström, 2001], confirman que un *tempo* rápido suele estar asociado a emociones como: felicidad, alegría, agrado, sorpresa, furia, inquietud, miedo. Expresiones que en determinado contexto se manifiestan como emociones fuertes. Mientras que un *tempo* lento se asocia a expresiones de calma, serenidad, tristeza, paz, aburrimiento, anhelo, etc.

### 2.3.2. Escalas

En los conceptos clave de la teoría musical, hablamos sobre las escalas, y un hecho interesante es notar que cuando se intenta enseñar a un joven músico a diferenciar entre una escala mayor de una escala menor, el maestro suele emplear una comparación emocional entre ambas, diciéndole a su alumno que las escalas mayores suenan felices, y las menores tristes. Así, gracias a los estudios realizados a lo largo de los años, se ha confirmado que las escalas mayores reflejan un sentimiento de alegría, agradecimiento y las escalas menores un sentimiento de tristeza o disgusto [Gabrielsson and Lindström, 2001].

### 2.3.3. Volumen o intensidad

Ponerle intensidad o fuerza a ciertas melodías de una pieza musical, nos permite darle más énfasis y emoción. La música fuerte se asocia con expresiones de intensidad, poder, tensión, ira, alegría y miedo. La música baja es asociada con expresiones más sumisas como tristeza, paz, calma [Gabrielsson and Lindström, 2001].

### 2.3.4. Octavas

Las octavas más agudas provocan un sentimiento de positivismo, mientras que las octavas más graves una sensación de negatividad [Gabrielsson and Lindström, 2001].

### 2.3.5. Armonías

Como mencionamos arriba, una armonía nos permite embellecer una canción, darle más cuerpo y expresión. Aquellas armonías que sean consonantes y placenteras para el oyente, se asocian con expresiones de felicidad, serenidad, soñadoras, y aquellas armonías que son más complejas y disonantes, expresan tensión, furia e inconformidad. Es interesante notar cómo una melodía disonante no denota siempre un conjunto de sonidos que suenan desagradables, sino que se utilizan para expresar emociones de más fuerza [Gabrielsson and Lindström, 2001].



### 2.3.6. Acordes

Se mencionó la existencia de varios tipos de Acordes, los acordes Mayores, expresa un sentimiento de felicidad, los menores de tristeza, los aumentados una sensación de tensión y los disminuidos sonidos de misterio. En los conceptos clave, no profundizamos en los acordes con séptima. Pero al agregarle a cada acorde su séptima se producen nuevas sensaciones [Rowe, 2001].

- Acordes mayores con 7ma Mayor producen sonidos cristalinos.
- Acordes mayores con 7ma menor sonidos de tensión
- Acordes disminuido con 7ma disminuida producen sonidos densos.
- Acordes con 7ma menor generan una sensación de paz y calma.

### 2.3.7. *Pitch* o tono

Hacer que una pieza cambie de tonalidad varias veces, provoca emociones como agradecimiento, felicidad, sorpresa. Cuando no hay muchos cambios en la tonalidad se producen sensaciones de disgusto, aburrimiento, miedo y enojo Gabrielsson and Lindström [2001].

## 2.4. Consideraciones finales

Hasta este punto se han presentado las bases de la teoría musical y la relación que poseen sus elementos para inducir emociones en el oyente. En el siguiente capítulo, profundizaremos sobre metodologías que hacen uso de estos recursos para la composición musical basada en emociones y los elementos visuales que existen para representar el contenido de una novela literaria.

## Capítulo 3

# Trabajos relacionados

A nuestro entender, no existen trabajos previos que combinen la composición musical automática basada en emociones y la generación de resúmenes visuales a partir de un texto al mismo tiempo. Por lo tanto, en esta sección se revisan por separado los trabajos relacionados con el análisis de emociones, la composición algorítmica y la generación de resúmenes visuales a partir de un texto.

### 3.1. Análisis de emociones en un texto

La detección de emociones en la lingüística computacional es el proceso de identificar emociones discretas expresadas en una pieza de texto. El análisis de emociones se puede ver como la evolución natural del análisis de sentimientos, en la cual nos limitamos a identificar si una porción de texto es positiva, negativa o neutra.

A diferencia de el análisis de sentimientos, el conjunto de datos para el análisis de emociones son insuficientes. Seyeditabari et al. [2018] menciona tres tipos de recursos para abordar este problema.

El primero de ellos son **Texto etiquetado**, aquí los datos son una serie de oraciones etiquetadas con su respectiva emoción, un ejemplo es el conjunto de datos de ISEAR, *International Survey On Emotion Antecedents*<sup>1</sup>, el cual contiene 7600 oraciones asociadas a los sentimientos de (alegría, miedo, enojo, tristeza, disgusto, vergüenza y culpa). Otro ejemplo sería el conjunto de datos de [Alm, 2008], estos datos incluyen oraciones anotadas de cuentos de hadas bajo cinco etiquetas (alegría, miedo, tristeza, sorpresa y enojo) de la lista de emociones básicas de Ekman. Encontramos también el conjunto de datos de Aman and Szpakowicz [2007, 2008]. Este conjunto de datos consta de oraciones llenas de emoción recopiladas de blogs. Y para el etiquetado también se consideraron las emociones básicas de Ekman (alegría, tristeza, enojo, disgusto, sorpresa y miedo) además de una categoría neutra.

El segundo recurso descrito, son los **Lexicones emocionales**. En contraste a los Textos etiquetados, los *lexicons* son un conjunto finito de palabras asociadas a una etiqueta específica, para este caso en particular, la etiqueta es una emoción. Aunque no existe una gran cantidad de *lexicons*, es posible nombrar algunos. Por ejemplo Valitutti [2004], poseen un

---

<sup>1</sup>Isear Dataset: [https://github.com/sinmaniphel/py\\_isear\\_dataset](https://github.com/sinmaniphel/py_isear_dataset)

conjunto de datos asociados a diversas emociones, clasificados como Positivos, Negativos, Neutros o Ambiguos y categorizados en 28 subcategorías (alegría, amor, miedo, etc.). Sin embargo, a pesar de su amplia variedad, el conjunto de palabras asociadas a una emoción es poco extensa. Otro conjunto emocional que se usa con frecuencia en la literatura es el *lexicon* LIWC James W. Pennebaker [2001], consta de 6400 palabras anotadas con su respectiva emoción. Finalmente encontramos al *National Research Council Canada Emotion Lexicon* o abreviado por sus siglas *NRC Emotion Lexicon*<sup>2</sup> de Mohammad and Turney [2013], que posee más de 14000 palabras y las asocia con ocho emociones básicas (ira, miedo, anticipación, confianza, sorpresa, tristeza, alegría y disgusto) y dos sentimientos (negativos y positivos). Recientemente [Krebs et al., 2018], han incrementado éste *lexicon* de 14,181 palabras a 31,485 buscando los sinónimos de cada palabra provista en el *lexicon* original y asociándolas al mismo vector de emociones. Los autores también proveen un conjunto de oraciones de *Facebook* etiquetadas con las mismas ocho emociones.

El tercer recurso son los **Word embeddings**, son un algoritmo de aprendizaje no supervisado para obtener representaciones vectoriales de palabras Seyeditabari et al. [2018]. Muchas veces no basta con representar una palabra con 0's y 1's como se hace en los *one-hot vectors*, sino que además es necesario saber el contexto en el que cada palabra se encuentra, es decir, cuál es su relación con respecto a las otras palabras dentro de la misma oración. Esto contribuye a identificar qué palabras son o no relevantes en una tarea específica. *Glove* Pennington et al. [2014] y *Word2Vector* Mikolov et al. [2013] son de los *embeddings* más populares en tareas de lenguaje natural.

En nuestros experimentos, estaremos utilizando tanto los *lexicons* emocionales como el texto etiquetado en combinación con los *Word Embeddings* para averiguar cuál de estas opciones se adapta mejor a nuestro modelo.

## 3.2. Composición algorítmica

La composición algorítmica o *Algorithmic Composition* (AC) por sus siglas en inglés, es la automatización parcial o total del proceso de composición musical mediante el uso de computadoras [Rodríguez and Vico, 2014].

En las siguientes cinco subsecciones procederemos a revisar los enfoques más comunes para la composición algorítmica que se emplean en el campo de *Artificial intelligence* (AI). Para una revisión más completa, revisar los trabajos de Rodríguez and Vico [2014], Williams et al. [2015]..

### 3.2.1. Enfoque basado en gramáticas

Una gramática formal es un conjunto de reglas para reescribir cadenas de caracteres. Por lo tanto, una gramática formal generalmente se piensa como una generadora de lenguajes [Hausser, 2014]. En nuestro caso un lenguaje musical.

Las gramáticas nos permiten representar sistemas con estructura jerárquica. Debido a que, las estructuras jerárquicas pueden reconocerse en la mayoría de los estilos de mú-

<sup>2</sup>NRC Emotion Lexicon: <https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm>

sica, no es de extrañar que la teoría gramatical formal se haya aplicado para analizar y componer música durante mucho tiempo [Rodríguez and Vico, 2014]. Así, el objetivo de este enfoque, es la creación de un lenguaje musical que permita la composición de piezas musicales, sin embargo, el problema con un enfoque gramatical es la dificultad para definir manualmente un conjunto de reglas gramaticales para producir buenas composiciones. También es necesario acotar que la creación de una gramática puede tardar mucho tiempo. Y a pesar que se han creado herramientas para la automatización de gramáticas [Jacob], otros todavía siguen diseñando sus gramáticas a mano, eligiendo cuidadosamente el mapeo entre símbolos terminales y objetos musicales.

### 3.2.2. Enfoque basado en reglas

Los *Sistemas-Basados en Reglas* (SBR) intentan representar el conocimiento como un conjunto símbolos más o menos estructurados. Y ya que el conocimiento sobre composición musical puede estructurarse como un conjunto de reglas más o menos formalizadas, los SBR son una forma natural de implementar la composición algorítmica [Rodríguez and Vico, 2014]. Dicho de otra forma, se usa un conjunto de reglas para crear estructuras musicales y esto implica que los autores del sistema dominen ampliamente los conocimientos sobre teoría musical ya que la calidad de las piezas musicales generadas por estos sistemas, dependerán de qué tan bien se transforme el conocimiento musical en reglas.

Algunos de los modelos bajo este enfoque son *Illiad Suite* de Hiller Jr and Isaacson [1957], el cual fue uno de los primeros sistemas basados en reglas para composición musical. *Robin* de Morreale et al. [2013], que se auto categoriza en este tipo de sistema y otro de ellos es *Robinflock* de Masu et al. [2017] que a diferencia del primero genera música polifónica.

Sobre este tipo de enfoques, podríamos decir que su punto más fuerte es también su punto más débil. Como se mencionó anteriormente, la calidad de la composición musical dependerá principalmente del conjunto de reglas sobre los que trabaje el sistema, lo que podría considerarse como una limitante si hablamos de sistemas autónomos, sin embargo, trabajar bajo este enfoque también involucra libertad para el autor en su composición.

### 3.2.3. Enfoque basado en las cadenas de Márkov

Las cadenas de Márkov o el modelo de Márkov es un tipo especial de proceso estocástico discreto en el que la probabilidad de que ocurra un evento depende solamente del evento inmediatamente anterior. Esta característica de falta de memoria recibe el nombre de propiedad de Markov [Basharin et al., 2004].

Las cadenas de Márkov pueden representarse como grafos o matrices y posee una gran variedad de aplicaciones entre ellas, composición musical. Cuando estas se aplican a la composición musical, las matrices de probabilidad pueden ser inducidas a partir de un corpus de composiciones preexistentes o derivadas a mano de la teoría musical [Rodríguez and Vico, 2014]. Por ejemplo podemos aplicar una de esas matrices es para determinar la progresión armónica. Se entiende progresión armónica como un flujo continuo de acordes.

Tabla 3.1: Matriz de Márkov para la selección de acordes. Extraído de [Morreale et al., 2013]

	I	II	III	IV	V	VI	VII
I	0	0.05	0.05	0.30	0.20	0.05	0.1
II	0.04	0	0.04	0.04	0.45	0.08	0
III	0	0.07	0	0.21	0.07	0.65	0
IV	0.15	0.10	0.05	0	0.35	0.05	0
V	0.64	0.05	0.05	0.13	0	0.13	0
VI	0	0.40	0.10	0.10	0	0	0
VII	0.8	0	0	0	0	0	0

El método consiste en elegir una escala de forma aleatoria y luego procesar iterativamente la matriz de Márkov para calcular los acordes sucesivos.

En el Cuadro 3.1 podemos observar una matriz de Márkov para la selección de escala y acordes.

### 3.2.4. Enfoque basado en redes neuronales artificiales

Las *Redes Neuronales Artificiales* (RNA) se utilizaron por primera vez durante los años setenta y ochenta para analizar composiciones musicales, pero luego fueron adaptados para composición musical [Rodríguez and Vico, 2014]. De entre todos los tipos de RNA, son las RNA Recurrentes las que han demostrado un mejor desempeño. El objetivo básico es recibir una secuencia de notas musicales y a partir de ellas generar y completar composiciones musicales de forma automática [García Torrecilla, 2018]. Como parte de los trabajos relacionados tenemos el trabajo de Chu et al. [2016], una red musicalmente plausible para la generación de música pop y el trabajo de Goel et al. [2014], que aplica una variación de RNA para la generación de música polifónica.

Actualmente este es uno de los enfoques más utilizados para la generación de música y a diferencia de el enfoque basado en reglas, las RNA intentan eliminar la intervención humana tanto como sea posible.

### 3.2.5. Enfoque evolutivo

Los algoritmos evolutivos son técnicas de optimización estocástica basadas en el proceso de evolución por selección natural [Morreale et al., 2013]. Su objetivo es desarrollar piezas musicales en el estilo de un compositor o género en particular. La desventaja de los modelos basados en un enfoque evolutivo reside en los resultados ya que la música generada por estos modelos tiende a sonar antinatural y experimental dando como resultado una pieza musical caótica y disonante. Artículos como los de Miranda and Al Biles [2007], Towsey et al. [2001] trabajan bajo este enfoque. El segundo de ellos emplea algoritmos genéticos para la composición musical y discute cómo construir una función de aptitud para la extensión melódica.

Además, al igual que el enfoque anterior, los algoritmos evolutivos también intentan au-

tomatizar el sistema de composición, sin embargo, esto provoca que carezca de estructura en su razonamiento, evitando que se desarrollen soluciones sutiles para resolver problemas de composición tales como la armonización [Morreale et al., 2013].

### 3.2.6. Algoritmos afectivos para composición

Adicionalmente a los cinco enfoques presentados anteriormente, decidimos hacer uso de esta sección para presentar los (AAC). Los AAC, es un campo que combina la composición asistida por una computadora y su evaluación emocional [Williams et al., 2015]. Según nuestra investigación, podemos deducir que es un campo relativamente nuevo, y que a diferencia de los enfoques anteriores no solo involucra conocimientos en teoría y composición musical sino también conocimientos en psicología musical, la cual nos permite entender la relación entre la estructura musical y las emociones [Williams et al., 2015]. Para nosotros existen diversas formas de abordar el campo de AAC, una de ellas es lograr que el algoritmo de composición musical refleje las emociones de una novela literaria, un cuento o un discurso. Existen trabajos relacionados a esta última definición como lo es el caso de *My Song* de Simon et al. [2008], que recibe como entrada la letra de una canción e intenta encontrar la mejor combinación de acordes que la acompañen.

Pero las aplicaciones que se vinculan estrechamente con nuestra propuesta son los modelos de Stere and Trausan-Matu [2017], que recibe como entrada una poema y emplean técnicas de AI para generar una melodía en piano como acompañamiento. De forma muy similar lo hacen los modelos de Davis and Mohammad [2014], Salas [2018].

Ambos generan piezas de música a partir de una novela literaria. Y aunque comparten el mismo enfoque las técnicas empleadas son distintas.

Además el modelo *Transprose* de Davis and Mohammad [2014] genera piezas en piano mientras que *Tambr* de Salas [2018] emplea múltiples sintetizadores.

## 3.3. Generación de resúmenes visuales a partir de un texto

En esta sección repasamos trabajos que relacionan las emociones o los significados presentes en el texto con metáforas visuales. Una propuesta de visualización de una obra literaria es presentada por [Yamada and Murai, 2009]. A través de la detección y el recuento de palabras clave en las obras de Shakespeare, se logra captar el significado de una historia, es decir, su patrón estructural. Además, crean una distribución visual de las palabras clave para extraer el contenido base de la historia. Para ello, el patrón estructural se expresa mediante una distribución escalar continua. Luego, la distribución numérica se convierte en una imagen a color que utiliza una paleta de colores original que es relevante para la historia. En este sentido, el texto se representa mediante imágenes visuales. Este mismo trabajo fue explorado más a fondo en Yamada et al. [2013]. Esta vez, los patrones de las historias en las novelas se visualizaron mediante el análisis de densidad. Se seleccionaron palabras clave en múltiples aspectos para tratar múltiples historias en paralelo. Una narrativa compleja puede ser visualizada simultáneamente con múltiples temas.

El trabajo presentado por [Mohammad, 2011] utiliza un conjunto de técnicas que accede automáticamente a los libros y los analiza (principalmente novelas y de los cuentos de hadas) utilizando el análisis de sentimientos en conjunto con visualizaciones efectivas. Esto permite cuantificar y rastrear las emociones tanto en libros individuales como en grandes colecciones. La visualización de emociones se realiza utilizando gráficos circulares, líneas de tiempo y nubes de palabras.

En nuestro enfoque, realizamos un resumen visual basado en un análisis topológico. Por lo tanto, los siguientes tres trabajos son los más relacionados con el nuestro. [Elson et al., 2010] muestran un método cuyo objetivo es extraer redes sociales de la literatura. La red se construye en base a los diálogos de los personajes de una novela. En este escenario, los personajes son los nodos del grafo y las aristas representan la una cantidad de conversaciones bilaterales entre dichos personajes. Los pesos de las aristas corresponden a la frecuencia y duración de sus conversaciones. Para ello, los autores proponer una nueva combinación de detección basada en patrones, estadística y la adaptación de las herramientas estándar de lenguaje natural para el género de alfabetización. Otro análisis topológico es realizado por [Waumans et al., 2015]. Construyen redes sociales basadas en el procesamiento de diálogos en el texto de las novelas. Se identifican los personajes que intervienen en cada conversación y se forma una red basada en estas interacciones. Se supone que la evolución temporal de la red sigue la sucesión de diálogos en los libros. El resultado final es la firma de cada historia, caracterizando elementos como el alcance, el número de protagonistas e incluso el estilo de escritura del autor. Por último, el trabajo de [Ardanuy and Sporleder, 2014] aborda dos enfoques para crear una red de personajes literarios, el primero, al igual que los dos anteriores, es una red conversacional en donde las aristas de la red representa una diálogo entre los personajes de la obra. El segundo enfoque, es una red de co-ocurrencias, en donde un nodo se relaciona con otro cuando dos personajes co-ocurren en una misma ventana de texto, como por ejemplo, un conjunto de párrafos en la obra. La intención de este método es abordar los trabajos literarios que carecen de diálogos. Luego, la red se construye en base dos abordajes: estático y dinámico. La red estática permite la representación visual de toda la novela. La red dinámica, es un conjunto de subredes que se construyen por cada capítulo de la obra. Finalmente, los autores deciden clasificar las redes resultantes por autor o género y descubren ciertos patrones que los hace similares.

## Capítulo 4

# Propuesta

El proceso para la elaboración de nuestro modelo consiste esencialmente en tres fases fundamentales: La primera de ellas es la extracción de emociones y personajes de obras literarias a través de algoritmos de procesamiento de lenguaje natural. La segunda es la composición musical basada en emociones, esta se conforma de tres módulos: Armonía, ritmo y melodía. La última fase de nuestro modelo es la animación del grafo. Finalmente, como resultado de la combinación de la melodía y el grafo animado, se obtiene el resumen literario. El *pipeline* de nuestra propuesta puede ser observado en la Figura 4.1. A continuación, describimos de forma más específica los pasos que nuestra propuesta involucra.

### 4.1. Extracción de emociones y personajes de obras literarias

Comenzamos nuestro *pipeline* con el preprocesamiento de la obra literaria. Nuestra metodología nos permite resumir cualquier trabajo literario en un formato de texto tradicional, por ejemplo, .pdf,.txt,.docx. Luego, realizamos los siguientes procedimientos:

1. Preprocesamiento del texto: El primer paso consiste en descartar información poco relevante del texto que se vaya a introducir en el modelo, para ello hacemos uso del *kit* de herramientas de lenguaje natural NLTK<sup>1</sup> y realizamos las siguientes subtarefas.

---

<sup>1</sup><http://www.nltk.org/>

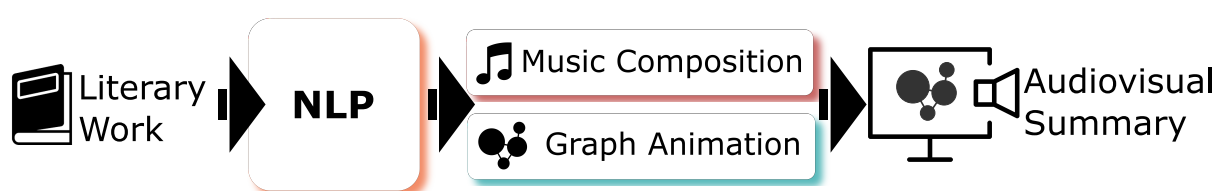


Figura 4.1: *Pipeline* empleado para nuestra metodología.



- 
- a) Remover cabeceras: Consiste en remover las cabeceras, títulos, subtítulos, tabla de contenidos, dedicatorias, glosario, etc. de la obra literaria.
  - b) Homogeneización de las comillas: Será necesario homogeneizar todo tipo de comillas a comillas dobles para denotar diálogos en la obra. Y usar comillas simples para denotar contracciones.
  - c) Contracciones: En el idioma inglés, las contracciones son muy frecuentes en su escritura. Por lo que pasar de **'s** a **is** o de **aren't** a **are not** también fue parte de nuestro preprocesamiento.
  - d) De mayúsculas a minúsculas: Esta conversión se logra rápidamente usando el comando **lower()** que pertenece a las funciones básicas del lenguaje de programación **Python**.
  - e) Remover *stopwords*: Decidimos remover los *stopwords* (the, is, are, etc.) haciendo uso de NLKT.
  - f) Remover símbolos de puntuación: A excepción de las comillas dobles y luego de haber preprocesado las contracciones. Eliminamos los símbolos de puntuación. Ninguno de ellos aporta información útil para la extracción de emociones y personajes.
  - g) Lematizar: El objetivo es transformar palabras a su raíz. Como por ejemplo, pasar de **running** a **run**. De esta forma, aumentamos la probabilidad de encontrar el verbo en el *lexicon*.
  - h) Itertools: Es común encontrar en los diálogos la deformación de ciertas palabras como **ooooooooohhh** para expresar intensidad. Es necesario convertir la palabra a su forma correcta, **oh**. Sin embargo, solo fue posible construir un conjunto limitado de las palabras más comunes en inglés que sufren esta alteración.
  - i) Conversor de abreviaciones: Las abreviaciones no son muy comunes en los trabajos literarios, pero a veces nos encontramos con algunas de ellas como **OMG** o **ASAP**. Por lo que es muy conveniente, intentar pasar estas expresiones a su formato más extenso como **Oh My God** y **As Soon As Possible**.
  - j) Convertir un verbo a su tiempo presente: Haciendo uso de la librería Nodebox<sup>2</sup>, es posible convertir un verbo en tiempo pasado o en tiempo futuro a su tiempo presente. Esto hace que sea mucho más fácil encontrar la palabra en el *lexicon*.
2. Conteo de emociones: En este paso, haremos uso de los dos recursos mencionados al final de la Sección 3.1 y en el Capítulo 5 analizaremos cuál de ellos nos brindó un mejor resultado.
- a) Lexicones emocionales: Como primer recurso, elegimos *Word-Emotion Association Lexicon* por su vasto número de palabras asociadas a una emoción. NRC nos ofrece las ocho emociones básicas propuestas por Robert Plutchik, de las cuales usaremos seis de ellas: alegría, tristeza, ira, miedo, anticipación y sorpresa, detalladas en la Tabla 4.1. Para nosotros, cuatro de estas emociones fueron categorizadas como primarias y las otras dos como secundarias. De esta manera, la alegría, la tristeza, la ira y el miedo son emociones primaria, mientras que la anticipación y la sorpresa son emociones secundarias. Luego, con el

---

<sup>2</sup><https://www.nodebox.net/>

texto ya preprocesado, se hace un recuento de las seis emociones básicas elegidas, además de los dos sentimientos (positivo o negativo) que el texto refleja. Este conteo, se realiza a lo largo de toda la obra y por un número predefinido de secciones. En nuestro caso, elegimos dividir el texto en cuatro secciones, pero el número de secciones puede depender enteramente de la longitud de la obra literaria.

- b) Texto etiquetado: Para emplear el segundo recurso, aplicaremos técnicas de aprendizaje profundo. El objetivo es que a partir de un conjunto de datos, podamos predecir la emoción asociada que mejor le corresponda. Para lograr esto, necesitaremos realizar los siguientes pasos:

- 1) Armar un *dataset*: En técnicas de aprendizaje profundo, tener un buen conjunto de datos es parte esencial del algoritmo. Como ya habíamos mencionado, una de las desventajas del análisis de emociones, es la escasez de datos. Para abordar este problema, decidimos aplicar la técnica de Chaffar and Inkpen [2011] y construir nuestro *dataset* combinando conjuntos de datos que compartan la misma clasificación de emociones (neutro, alegría, tristeza, enojo, miedo). De esta forma, nuestro *dataset* global, se encuentra conformado por los siguientes sub-conjuntos de datos: *Emotion Intensities in Tweets* de [Mohammad and Bravo-Marquez, 2017], que proporciona un conjuntos de datos de entrenamiento y pruebas para cuatro emociones: alegría, tristeza, miedo y enojo. También combinamos los datos etiquetados por [Alm, 2008], que asocia las oraciones de decenas de cuentos de hadas con su respectiva emoción (Ira-Disgusto, Miedo, Alegría, Tristeza, Sorpresa y Neutro). Por último, empleamos el conjunto de emociones *Isear*, que reporta situaciones en las que los encuestados hayan experimentado alguna de las siguientes siete emociones (alegría, miedo, enojo, tristeza, disgusto, vergüenza y culpa). Al combinar todos estos datos, uno de los requisitos más complejos fue intentar mantener el conjunto lo más balanceado posible.
- 2) Limpiar el *dataset*: Este objetivo lo alcanzamos aplicando los pasos del preprocesamiento que se empleo en la Sección 4.1.
- 3) Escoger un modelo para la predicción de emociones: Decidimos escoger una Red Neuronal Recurrente (RNN), específicamente una *Long-Short Term Memory* (LSTM) combinada con una *Convolutional Neural Network* (CNN) tal y como se aplica en Sosa [2017]. Las LSTM, son un tipo de RNN cuya arquitectura está diseñada para recordar los valores leídos previamente durante un período de tiempo dado y son sumamente utilizados en tareas de procesamiento de lenguaje natural. De manera intuitiva, la ventaja de usar LSTM cuando se realiza cualquier tipo de análisis de texto es que la red recordará lo que leyó anteriormente y, por lo tanto, puede tener una mejor comprensión de la entrada. Por otro lado, aunque las CNN suelen aplicarse con mayor frecuencia en reconocimiento de imágenes, también es posible aplicar su arquitectura a tareas de lenguaje natural. Así, la red podrá extraer las características más relevantes independientemente de dónde se encuentren en la oración. Finalmente, el modelo LSTM-CNN consiste en una capa LSTM inicial que recibirá las palabras de cada oración representadas por los *embeddings* como entradas y la intuición es que sus *tokens* de salida almacenarán información no solo

del *token* inicial, sino también de todos los *tokens* de la oración. La salida de la capa LSTM sirve de entrada a una capa de convolución que extraerá las características locales más relevantes.

Finalmente, usaremos una red completamente conectada que tendrá cinco salidas, una para cada emoción (neutro, alegría, tristeza, enojo, miedo). Adicionalmente, hicimos que la primera capa de LSTM se compusiese de neuronas LSTM bidireccionales<sup>3</sup>, las cuales se basan en la idea de que la salida en el tiempo *t* puede depender no solo de los elementos anteriores en la secuencia, sino también de los elementos futuros.

- 4) Conteo de Emociones: Una vez finalizado el modelo y procesado el texto. Será necesario contar las emociones con las que se asoció a cada oración.
3. Cálculo de densidades: A veces será necesario calcular la densidad de una emoción específica. Para ello, contamos cuántas veces una palabra se relaciona con dicha emoción en la sección *Si* y lo dividimos por la cantidad de palabras en total que hay en la sección *Si*. Si en cambio, se solicita la densidad de una emoción en toda la obra. La densidad es la cantidad de veces que una palabra se relaciona con dicha emoción en **toda la novela** sobre la cantidad de palabras en **toda la novela**.
  4. Identificación de personajes principales y diálogos: Nuestro método extrae tanto personajes como diálogos para que luego sean empleados en nuestra representación visual, que se detalla en la Sección 4.3. Para ello, realizamos los siguientes pasos:
    - a) Hacemos uso de la librería Spacy<sup>4</sup> para identificar a los personajes en la obra literaria.
    - b) Elaboramos una lista de los títulos más comunes en inglés como: Sir, Mr, Miss, Lady, Lord, etc.
    - c) También, elaboramos una lista de los nombres más comunes en obras literarias, en caso que el módulo de Spacy no los detecte.
    - d) En seguida, hacemos uso del módulo Nameparser<sup>5</sup>, para reconocer los componentes individuales de un nombre y registrar las entidades de forma más estructurada.
      - 1) Título + Nombre + Apellido
      - 2) Título + Apellido
      - 3) Título + Nombre
      - 4) Nombre + Apellido
      - 5) Nombre o Apellido

Para este procedimiento será necesario corroborar que el género del nombre corresponda con el título que lo antecede.
    - e) Luego desarrollamos un algoritmo que nos permite agrupar las correlaciones de los personajes. Por ejemplo, Wendy es el mismo personaje que Miss Wendy, Wendy Darling, Miss Darling, etc. Así, en el grafo sólo una de las correlaciones se muestra.

<sup>3</sup>Recurrent Neural Networks Tutorial: <http://www.wildml.com>

<sup>4</sup><https://spacy.io/>

<sup>5</sup><https://pypi.org/project/nameparser/>

Tabla 4.1: Emociones básicas y avanzadas. Basado en Plutchik [1984].

Emoción básica	Emoción avanzada	Combinación
Alegría	Optimismo	Alegría + Anticipación
Confianza	Amor	Confianza + Alegría
Miedo	Sumisión	Miedo + Confianza
Sorpresa	Susto	Sorpresa + Miedo
Tristeza	Decepción	Tristeza + Sorpresa
Aversión	Remordimiento	Aversión + Tristeza
Ira	Desprecio	Ira + Aversión
Anticipación	Alevosía	Anticipación + Ira

- f) En casos donde se presenten dos o más personajes con el mismo título como Miss Darling para introducir a los personajes de Wendy y la Madre de Wendy en la obra “*Peter Pan and Wendy*”. Optamos por hacer caso omiso a esta correlación.
- g) A continuación, identificamos los diálogos de la obra, reconociendo el texto dentro de cada par de comillas dobles. Más adelante, los diálogos que corresponden a la misma conversación en la novela deben ser agrupados. Para una revisión más específica de este procedimiento verificar Waumans et al. [2015].
- h) Finalmente, tendremos una tabla que registre cada nueva interacción entre dos personajes específicos y realizaremos un conteo de la cantidad de veces que un personaje aparece en un diálogo. Esta información será utilizada más adelante para la animación del grafo.

## 4.2. Composición musical basada en emociones

La segunda fase de nuestro *pipeline* es la Composición musical basada en emociones, esta etapa engloba todos los módulos necesarios para la composición de una melodía con múltiples instrumentos que se basa en la información recibida de la primera etapa, la cual se resume en: La cantidad de palabras asociadas a una emoción principal alegría, tristeza, ira, miedo, y la cantidad de palabras asociadas a una emoción secundaria como anticipación y sorpresa (Confianza y Disgusto no se tomaron en cuenta en este trabajo). Además, se obtuvo el número de palabras asociadas a un sentimiento positivo y negativo. Y el número de palabras en total (sin tomar en cuenta aquellas que fueron eliminadas por ser *stopwords*). Este conteo, se realizó por cada sección de la novela (Si). A continuación describimos los tres módulos que componen la melodía.

### 4.2.1. Módulo de la armonía

El objetivo de este módulo consiste en elegir una progresión de acordes que mejor se adapte a cada emoción. La progresión de acordes es una parte sumamente importante

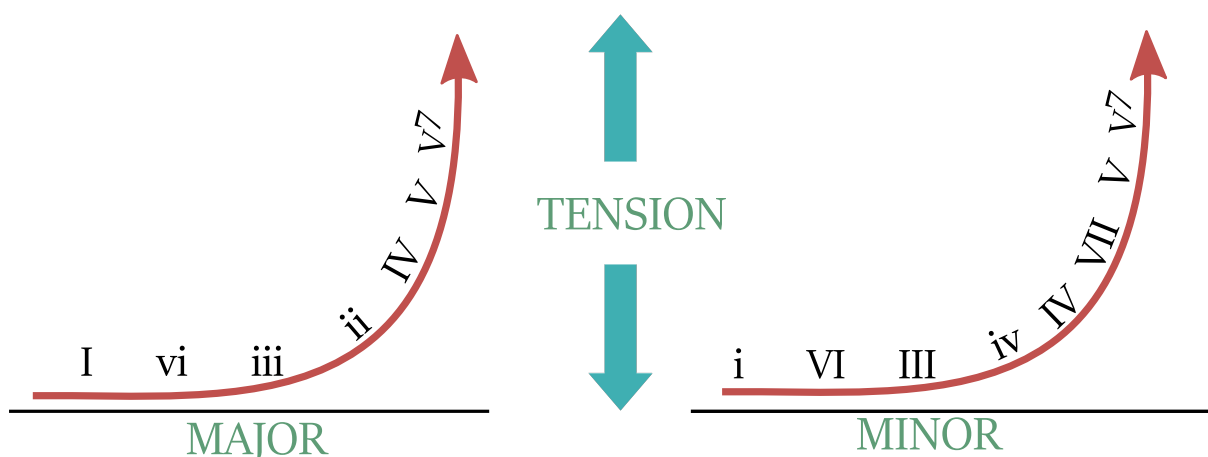


Figura 4.2: Acordes más populares para escalas mayores y menores según [Findeisen, 2015]

en la música, con ella podemos generar tensión y resolución en una melodía. Para este algoritmo, nos basamos en los acordes más populares descritos en Findeisen [2015]. En la Figura 4.2, se muestran una serie de acordes para escalas mayores y menores. Según el orden en que se dispongan estos acordes, es posible generar mucha o poca tensión. Entonces, los acordes que se encuentran más cerca al inicio de la curva (I, vi, i, VI), son los que menos tensión generan, y mientras más lejos se escoja el siguiente acorde con respecto al inicio de la curva, más tensión habrá (V, V7). Por ejemplo:

- I - vi tiene muy poca tensión.
- vi - V tiene mucha tensión.

De esta forma, para asignar una progresión de acordes a cada  $S_i$ , hallamos la distancia de la emoción más predominante en  $S_i$  hacia la emoción más predominante en  $S_{i+1}$ . Formándose tres líneas con pendientes positivas, negativas o con pendiente cero como se observa en la Figura 4.3.

- Si la pendiente es negativa o igual a cero, elegiremos una progresión de acordes que genera alta tensión y luego iremos añadiendo otros acordes que vayan disminuyendo la tensión.
- Si en cambio, la pendiente es positiva, se empezará con acordes de baja a alta tensión.

Este proceso se realizará para las tres primeras secciones de la novela y únicamente en la última sección ( $S_4$ ), se usará una progresión de acordes clásica como (I-IV-V-I). Una vez elegida una progresión de acordes para cada sección, se añadirá color a las notas para acentuar las emociones. Por ejemplo: Si la emoción predominante en  $S_i$  es alegría, puede elegirse aumentar los acordes mayores y se debe elegir la mayor cantidad posible de acordes mayores de la escala. Para tristeza, pueden disminuirse los acordes menores e intentar elegir la mayor cantidad posible de acordes menores de la escala. Para ira, puede añadirse una séptima mayor a los acordes mayores y para miedo, puede añadirse una séptima menor a los acordes menores. Para los casos de anticipación y sorpresa,

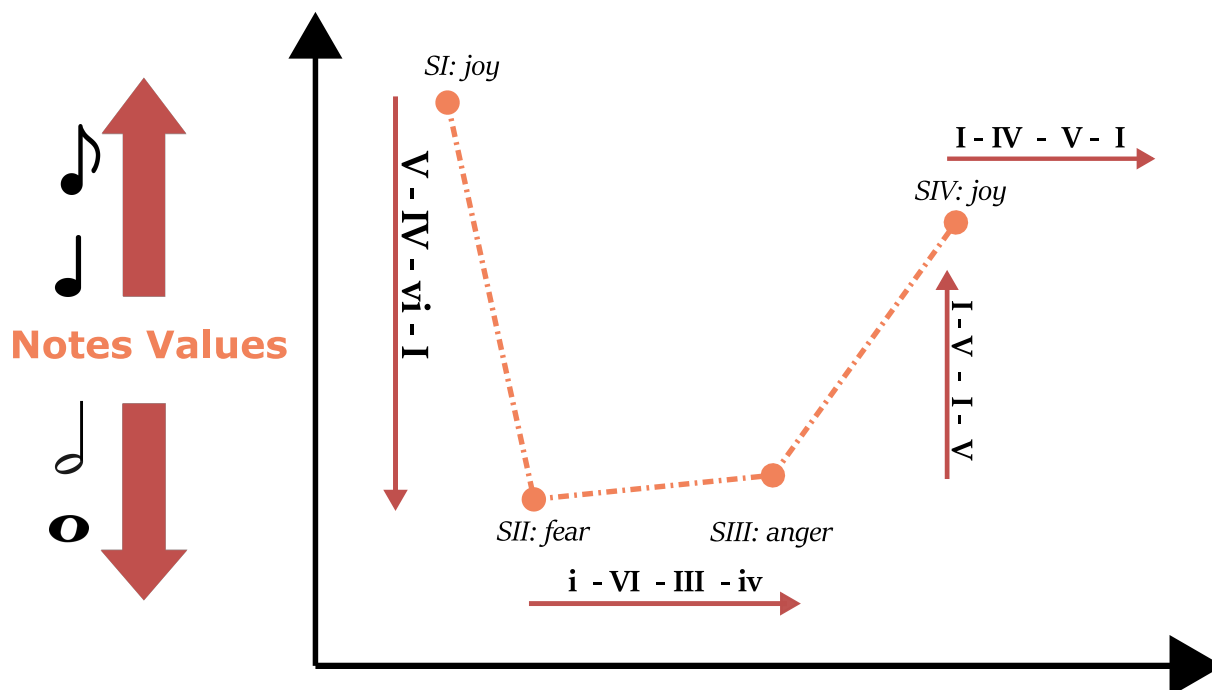


Figura 4.3: Progresiones de acordes seleccionadas para la obra “*Peter Pan and Wendy*”.

emociones secundarias, se añade color con respecto a la emoción primaria de toda la novela. Un ejemplo de esta selección de acordes puede encontrarse en la Figura 4.3.

#### 4.2.2. Módulo del ritmo

El módulo del Ritmo se compone de tres pasos esenciales:

1. Lo primero a realizar es la selección del Tempo, para calcularlo emplearemos la fórmula de Davis and Mohammad [2014] con la diferencia de que el tempo variará entre 60 y 180 bpm.

$$tempo = 60 + \frac{(Act - Actmin) * (180 - 60)}{Actmax - Actmin} \quad (4.1)$$

La ecuación 4.1 describe que a más actividad en el texto en general, más tempo habrá. De esta forma el término *Act*, es la diferencia de las densidades activas y pasivas, siendo la alegría y la ira emociones activas y la tristeza una emoción pasiva. Recordemos por último que por densidad nos referimos al conteo de una emoción específica sobre la cantidad de palabras en total.

2. El segundo paso, consiste en seleccionar el *Time Signature*, para este modelo, nos limitaremos a un compás de 4/4 que es el más común.
3. El paso final consiste en generar el patrón de ritmo, Para ello, nuevamente haremos referencia a la Figura 4.3. Como podemos observar, hacemos uso de cuatro posibles duraciones (blancas, negras, medias y corcheas). Mientras más larga sea la distancia de una sección a otra, mayor variedad de duración de notas se abarca. Por ejemplo,

la primera distancia de la obra "*Peter Pan and Wendy*", parte desde el punto más alto hasta el punto más bajo de la obra, por lo que el ritmo podrá ser el resultado de combinar notas con duración más corta (corcheas) hasta notas desde la duración más larga (blancas). Mientras que de la distancia de la sección  $S_2$  a  $S_3$ , solo dispondrá de notas blancas, puesto que es la distancia más corta de la obra y abarca menos duraciones. La distribución de las duraciones de las notas, se logra dividiendo la distancia del punto más alto al punto más bajo por cuatro y ubicando las duraciones entre cada sector tal y como se observa en la Figura 4.3. Además si la pendiente es positiva, las notas se ordenarán de forma creciente, caso contrario se ordenarán de forma decreciente. Fortaleciendo la sensación de dirección del ritmo.

Únicamente en la última sección ( $S_4$ ), podrá elegirse cualquier duración de notas entre (blancas, negras, medias y corcheas) y podrán ser dispuestas en cualquier orden. Finalmente, nuestro patrón rítmico se compone de cuatro compases, por lo que la repetición de cada uno de ellos será necesaria para que suene menos cohesivo. De este modo, se obtiene un total de ocho compases por sección. El número de compases puede variar en relación a la longitud de la obra literaria.

**Algorithm 1:** Selección de escala según la emoción predominante

---

```

Input:  $eS, eN$ 
Output:  $keys$ 
//  $eS$  representa la emoción más predominante en  $Si$ .
//  $eN$  representa la emoción más predominante en toda la novela.
//  $keys$  retornará el valor de la escala elegida.
1 if  $eS$  in  $[alegría, tristeza, miedo, ira]$  then
2   if  $eS == alegría$  then
3      $keys = [Cmaj, Gmaj]$ 
4   else if  $eS == tristeza$  then
5      $keys = [Cmin, C\#min, Bmin]$ 
6   else if  $eS == miedo$  then
7      $keys = [D\#min, Fmin, C\#min]$ 
8   else if  $eS == ira$  then
9      $keys = [Abmin, Fmaj, F\#min]$ 
10 else
11   if  $eS == anticipation$  then
12     if  $eN == alegría$  then
13        $keys = [Dmaj, F\#maj, Bbmaj]$ 
14     else if  $eN == tristeza$  then
15        $keys = [Dmin, F\#min, Bbmin]$ 
16     else if  $eN == miedo$  then
17        $keys = [D\#min, Bbmin]$ 
18     else if  $eN == ira$  then
19        $keys = [Fmaj, Abmaj]$ 
20   else if  $eS == surprise$  then
21     if  $eN == alegría$  then
22        $keys = [Amaj, Bbmaj]$ 
23     else if  $eN == tristeza$  then
24        $keys = [Emin, Fmin]$ 
25     else if  $eN == miedo$  then
26        $keys = [Fmin, Bbmin, C\#min]$ 
27     else if  $eN == ira$  then
28        $keys = [Gmin, Bmaj]$ 
29 return  $random(keys)$ 

```

---

**4.2.3. Módulo de la melodía**

Nuestro último módulo consiste en la selección de la escala y la octav además de asignar las notas musicales a nuestro patrón rítmico. La escala se elige en función de la emoción predominante en cada sección  $Si$ . Es importante tener en cuenta lo siguiente: si



la emoción con mayor predominancia pertenece a una de las emociones primarias (alegría, tristeza, ira, miedo), la selección de la escala se hará en función a esta emoción, sin embargo, si la emoción que predomina es una emoción secundaria (sorpresa o anticipación), será necesario encontrar la emoción primaria que predomina a lo largo de la novela ( $eN$ ), por lo que, para esos casos, la escala será seleccionada en base a ambas emociones. Adjuntamos el *Algorithm 1* para una mejor comprensión. El parámetro  $eS$  se refiere a la emoción predominante en  $S_i$ .

La forma de asignar las escalas para cada emoción se basó en Chase [2006], Hobbs [2018]. A continuación, tenemos que definir en qué octava del piano se tocará la melodía. Para ello, utilizamos el algoritmo propuesto por [Davis and Mohammad, 2014].

$$Octava = 4 + \frac{(JS - JSmin) * (6 - 4)}{JSmax - JSmin} \quad (4.2)$$

El término  $JS$  representa la diferencia entre las densidades de alegría y tristeza en cada  $S_i$ . Luego aplicamos la ecuación 4.3 basada en Davis and Mohammad [2014], donde ahora, al valor obtenido en  $Octava$ , le añadimos uno si la emoción predominante  $eS$  es alegría, disminuimos uno si la emoción predominante  $eS$  es ira, añadimos uno si la emoción predominante  $eS$  es miedo o tristeza y el número de palabras positivas en  $S_i$  es mayor que el número de palabras negativas en  $S_i$ , disminuimos uno si la emoción predominante  $eS$  es miedo o tristeza y el número de palabras negativas en  $S_i$  es mayor que el número de palabras positivas en  $S_i$ . Si las emociones son sorpresa o anticipación, el valor de  $Octava$  no se modifica.

$$Octava = \begin{cases} Octava + 1 & eS = \text{alegría} \\ Octava - 1 & eS = \text{ira} \\ Octava \pm 1 & eS = \text{miedo} = \text{tristeza} \\ Octava & \text{otherwise} \end{cases} \quad (4.3)$$

Luego, asignamos notas a nuestro patrón rítmico. La selección de las notas debe permitir que la melodía suene consonante para emociones como alegría o tristeza y disonante para emociones como miedo o ira.

- Para lograr una mayor consonancia, es necesario elegir la mayoría de las notas del acorde seleccionado y elegir las notas restantes de la escala seleccionada.
- Si quieres generar tensión, normalmente se eligen notas fuera del acorde y de la escala.

Para acentuar la emoción predominante en cada melodía, se añadieron algunos adornos al momento de construir la melodía como: floreos o notas de paso, ofreciendo una mayor variedad melódica. También se tomó en cuenta el posicionamiento de las notas. Si en la sección  $S_i$ , había mayor cantidad de palabras positivas, entonces las notas se elegían en dirección descendente de la escala. Si había más palabras negativas, la dirección de las notas era ascendente.

## Selección de instrumentos

Finalmente, para acentuar la emoción predominante en cada melodía, se realizó una lista de instrumentos, ofrecidos por la API de JFugue, que nos ayudaron a añadir color. Fueron elegidos de acuerdo a los criterios descritos en la Tabla 4.2. Como resultado, el archivo .midi generado por la API de JFugue reproduce la melodía bajo la estructura de una fuga musical. Por esta razón, los instrumentos elegidos fueron insertados y retirados a lo largo de toda la melodía. En algunos casos, se generó una segunda melodía, como acompañamiento, basada en la segunda emoción predominante en cada  $S_i$ .

Tabla 4.2: Lista de instrumentos de la API JFugue asociados a cada emoción

Emoción	Instrumentos
Alegría	Piano, Guitar, Flute, Piccolo, Marimba, Music Box
Tristeza	Piano, Glockenspiel, Orchestral Strings, Celesta
Ira	Timpani, Trombone, Viola, Cello, Contra-bass
Miedo	Violin, Reverse Cymbal, Vibraphone
Sorpresa	Violin, Piano.
Anticipación	Piano, Flute, Vibraphone

## 4.3. Animación del grafo

Con el fin de complementar el conjunto de conocimientos que proporciona nuestra metodología, hemos decidido introducir un recurso visual. Por lo tanto, implementamos una animación basada en grafos para representar cómo interactúan los personajes entre ellos, y así ir descubriendo quiénes son los más nombrados en la obra literaria, y con quién se relaciona más. Como se comentó en la Sección 4.1, nuestro paso de preprocesamiento extrae los diálogos y los personajes. Utilizamos esta información para mostrar la variación de frecuencia de cada personaje a lo largo de la historia utilizando los nodos del grafo, cada vez que un personaje dialoga con otro, se incrementa el tamaño de su nodo. Las relaciones entre los personajes están representadas por los aristas de los nodos, donde el ancho de la arista representa la frecuencia de interacción entre dos personajes específicos, es decir, cuanto mayor sea el número de diálogos entre dos personajes en la obra literaria, mayor será el ancho del arista que une sus dos nodos.

En nuestra implementación, utilizamos el *software* de visualización Graphviz<sup>6</sup> para la generación automática de grafos por su buen desempeño en términos de tiempo de procesamiento y flexibilidad para el dibujo de gráficos, y D3.js<sup>7</sup> para la inserción de grafos en el contenido de una página web y para la generación de transiciones. Además, introducimos una serie de restricciones para producir una representación legible y estética del gráfico, por ejemplo: la eliminación de superposiciones, la disposición jerárquica y la proximidad entre nodos, utilizando la herramienta DOT.

<sup>6</sup>[www.graphviz.org/](http://www.graphviz.org/)

<sup>7</sup>[www.d3js.org](http://www.d3js.org)

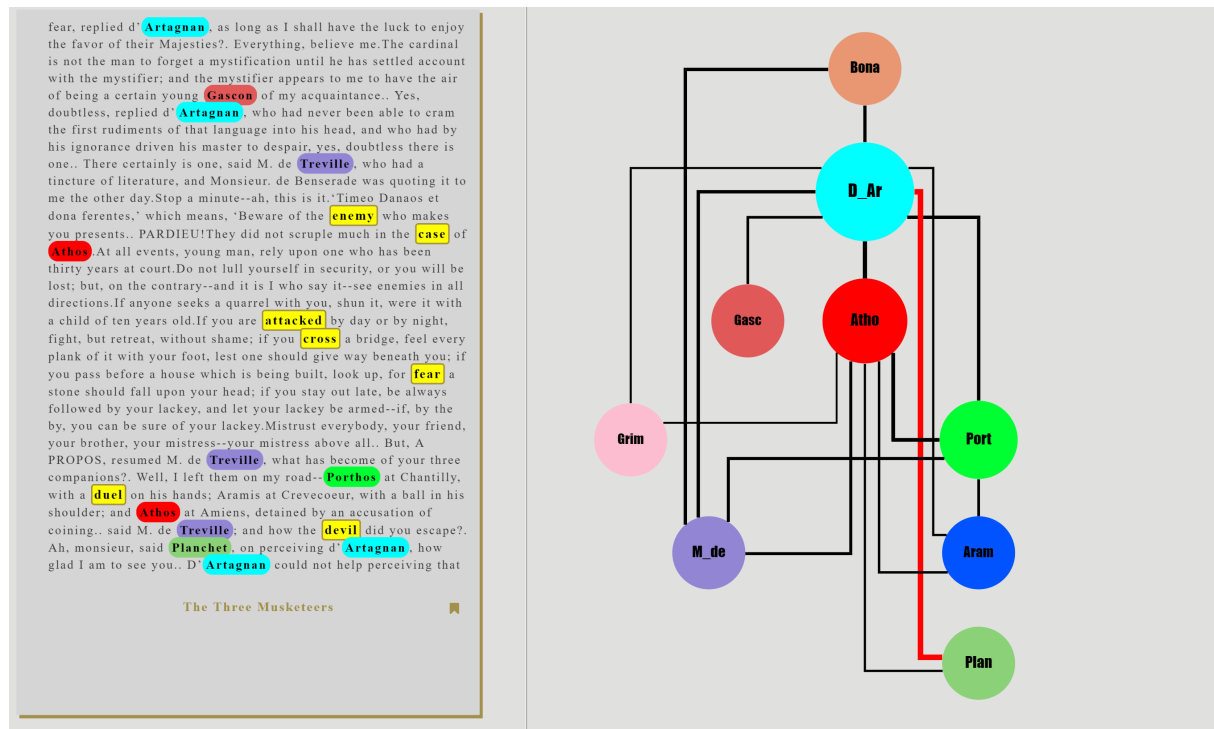


Figura 4.4: Una visión general de nuestro prototipo: (izquierda) una versión resaltada de los diálogos de texto y (derecha) un grafo animado que ilustra la interacción entre personajes.

En la Figura 4.4, mostramos una visión general de nuestro prototipo. A la derecha, el grafo animado para ilustrar las interacciones entre los personajes. En el lado izquierdo, mostramos los diálogos en detalle. Los colores se conservan en ambos lados de los prototipos para identificar rápidamente a los personajes. Además, destacamos (en amarillo claro) las principales palabras/acciones que los diálogos dicen. Esta simple característica ayuda al usuario a inferir perceptivamente la razón que relaciona a dos o más personajes.

## Capítulo 5

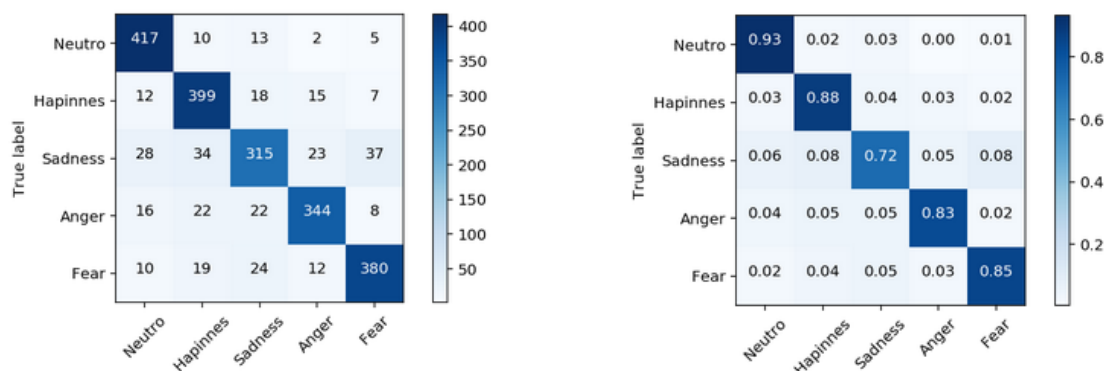
# Resultados

En esta sección, presentamos los resultados producidos por nuestra metodología. Seleccionamos un subconjunto de cinco *best sellers* de la literatura como: “*Peter Pan and Wendy*” de James Barrie, “*Psycho*” de Robert Bloch, “*The Lovely Bones*” de Alice Sebold, “*Don Quixote of La Mancha*” de Miguel de Cervantes y “*The Three Musketeers*” de Alexandre Dumas, para evaluar su capacidad de síntesis.

### 5.1. Conteo de emociones usando los recursos de: Texto etiquetado y Lexicones emocionales

Para la extracción y conteo de emociones, realizamos múltiples pruebas con dos de los recursos más utilizados en el estado del arte. Estos fueron profundizados en la Sección 4.1.

Para hacer uso del recurso **Texto etiquetado**, contruímos un *dataset* homogéneo con datos que estuviesen clasificados con la misma etiqueta (Neutro, Happiness, Sadness, Anger, Fear). De esta forma, nuestro conjunto de datos se encuentra compuesto por: diálogos de cuentos de hadas, comentarios de la Red Social *Twitter*, situaciones de la vida cotidiana, entre otros. Para comprobar la precisión de la red **LSTM-CNN** que se aplicó al conjunto, realizamos pruebas de *precision*, *recall* y *f1-score*. En la Figura 5.1, se puede observar la matriz de confusión que evalúa la capacidad de predicción de la red neuronal en un sub-conjunto de datos de entrenamiento. A primera vista, se puede notar que pese a la escasez de datos, se intentó balancear el conjunto para mejorar su resultado. Y aunque se llegan a alcanzar buenos porcentajes de *precision* para etiquetas como *Neutro* (93%) y *Happiness* (88%). El resultado de haber probado la red en un conjunto no etiquetado (los cinco *best sellers* mencionados anteriormente) prueban que los datos son simplemente insuficientes. Esto se debe principalmente a que la red tiende a etiquetar cada oración de la obra literaria como *Happiness*, consecuentemente, obras como “*Psycho*” (Figura 5.2b), cuyo tono es sumamente oscuro, fueron clasificados como alegres. De igual forma, es posible observar un patrón en la predicción del recurso, ya que en las tres obras procesadas en la Figura 5.2, las dos emociones más predominantes son *Happiness* y *Anger*, que a su vez, son emociones con un porcentaje de *precision* mayor



(a) Matriz de confusión, sin normalizar.

(b) Matriz de confusión, normalizada.

Figura 5.1: Matriz de confusión que evalúa el conjunto de datos homogéneo.

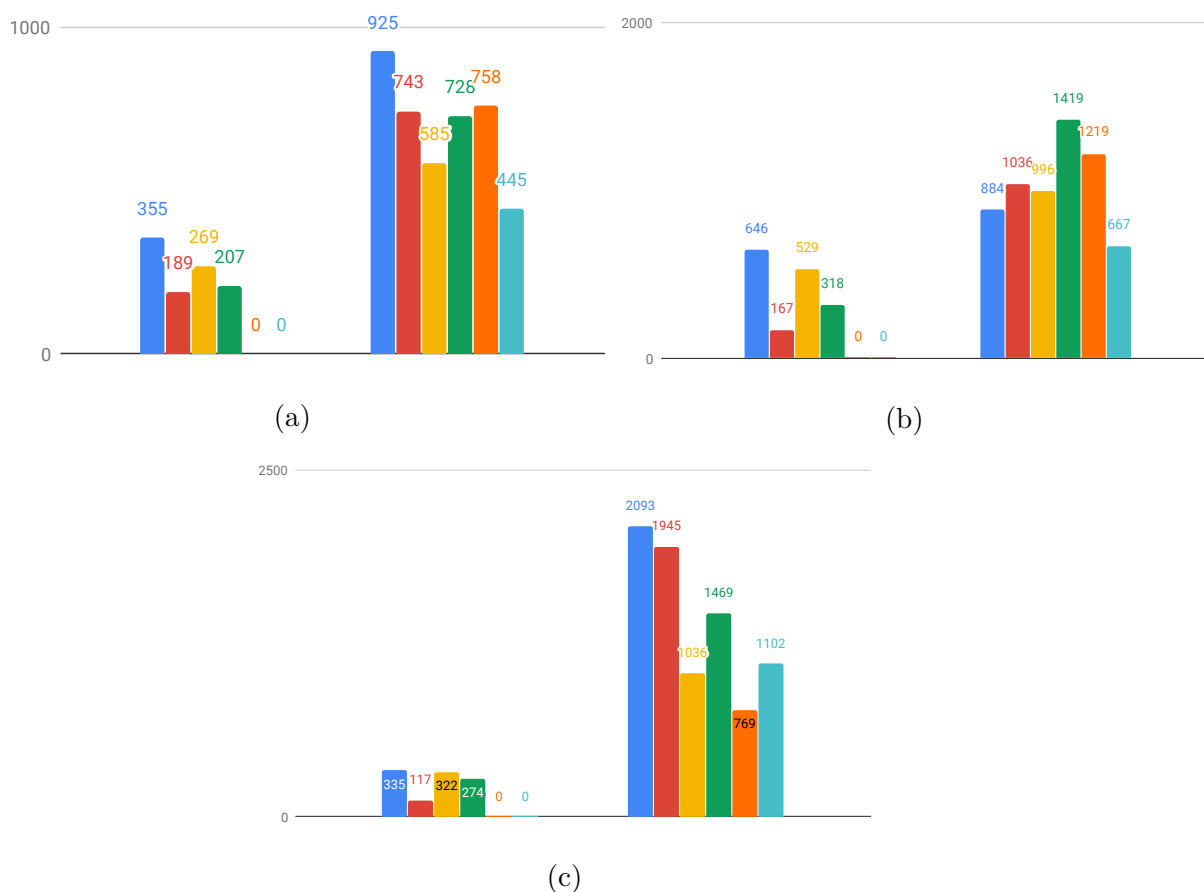


Figura 5.2: Extracción de emociones usando dos recursos diferentes: Texto etiquetado (izquierda) y Lexicones emocionales (derecha). Aplicado a las siguientes obras: “*Peter Pan and Wendy*”, “*Psycho*” y “*The Lovely Bones*” respectivamente. Los colores representan las emociones reconocidas: ■ = alegría, ■ = tristeza, ■ = ira, ■ = miedo, ■ = anticipación, y ■ = sorpresa.

al (80 %) en la matriz de confusión. Por lo que podemos deducir la existencia de un des-balance en datos.

Luego decidimos aplicar el recurso de **Lexicones emocionales**, por lo que hicimos uso de NRC Lexicon (con más de 14000 palabras asociadas a 7 emociones). Después de haber realizado el pre-procesamiento descrito en la Sección 4.1, se realizó el conteo de emociones con el *lexicon*. En la Figura 5.2, podemos observar tres diferencias marcadas con respecto al recurso precedente. Primero, empleando el recurso anterior (LSTM-CNN), se detectaron menos palabras asociadas a una emoción. El ejemplo más evidente de este aspecto puede diferenciarse en 5.2c, donde para la obra “*The Lovely Bones*” se detectaron solo 117 palabras asociadas a la emoción tristeza contra las 1945 palabras identificadas con la misma emoción por el *lexicon*. La segunda diferencia es la cantidad de emociones, con el conjunto de datos homogéneo que se construyó para la red descrita, se utilizaron datos asociados a solo 4 emociones (sin contar Neutro), mientras que con NRC es posible detectar 7 emociones. La última y tercer diferencia es la calidad en la predicción de emociones en una obra literaria. En ambos recursos ocurrió la siguiente premisa: *A mayor cantidad de texto, mejor predicción de emociones*. Sin embargo, debido a la escasez de datos del primer recurso empleado, la emoción con mayor predominancia en todas las obras procesadas fue *Happiness*, mientras que con el segundo recurso, obtuvimos una clasificación más apegada al género de la obra. Para “*Peter Pan and Wendy*” se etiquetó a toda la obra como alegre, para “*Psycho*” de Robert Bloch, se la clasificó a toda la obra con la emoción de miedo y para “*The Lovely Bones*”, se eligió alegría seguida de tristeza como sus emociones más predominantes. A pesar de ello, es necesario acotar que esta precisión no ocurre en textos cortos como canciones, poemas, etc.

En conclusión, de elegir el recurso de **Texto etiquetado**, terminaríamos por sacrificar la cantidad de emociones, la precisión de las emociones predominantes en la obra y sobre todo el tiempo que se consume en la parte de entrenamiento de la red neuronal. Por ello, para este trabajo decidimos elegir el recurso de **Lexicones emocionales** acompañado de sus ventajas y desventajas.

## 5.2. Extracción de características, grafos y pentagramas

Luego de haber elegido NRC Lexicon para el conteo de emociones, fue necesario realizar la extracción de las características más relevantes para la composición como lo son: el valor *Act* para hallar el tempo, el valor *JS* para hallar las octavas y las emociones más predominantes en cada sección (*Si*) de las obras procesadas. Así en la Tabla 5.1 se detallan estos valores, incluidas las emociones  $e_1$  y  $e_2$ . Como se puede observar en una simple inspección, la mayoría de los valores obtenidos son los esperados debido al tipo de libro, por ejemplo, alegría y anticipación para “*Peter Pan and Wendy*”, anticipación y sorpresa en “*The Lovely Bones*” y alegría y miedo en “*The Three Musketeers*”. Analizando el valor *JS* por cada libro, podemos observar que en la mayoría de los casos este valor está por encima de cero, lo que significa que en general las emociones extraídas son positivas, excepto en el caso específico de “*Psycho*” en donde los valores presentes son muy cercanos o por debajo de cero, lo que refleja las emociones negativas identificadas. Dicha característica también ocurre en algunas secciones del resto de los libros, específicamente en  $S_2$  de “*Peter Pan and Wendy*” y  $S_4$  de “*The Three Musketeers*”. Un aspecto interesante es el de los valores de la actividad (*Act*), ya que los valores de rango (máximo menos mínimo) entre las secciones de la misma obra literaria es de alrededor de 0.01, sin embargo, el texto de

Tabla 5.1: Emociones extraídas, actividad, tempo, octavas por sección en cada obra literaria procesada.

	" <i>Peter Pan and Wendy</i> "	" <i>Psycho</i> "	" <i>The Lovely Bones</i> "	" <i>Don Quixote of La Mancha</i> "	" <i>The Three Musketeers</i> "
$S_1$ -Act	0.058	0.029	0.032	0.057	0.043
$S_1$ -JS	0.0313	-0.0055	0.0051	0.0129	0.0082
$S_1$ -Tempo	180	88	98	177	120
$S_1$ - $e_1$ - $e_2$	alegría-anticipación	anticipación-miedo	anticipación-alegría	anticipación-alegría	anticipación-alegría
$S_1$ -Octavas	7 - 6	5 - 6	5 - 6	6-6	6-6
$S_2$ -Act	0.035	0.025	0.03	0.057	0.042
$S_2$ -JS	-0.0051	-0.0056	0.005	0.0113	0.0097
$S_2$ -Tempo	107	76	92	177	133
$S_2$ - $e_1$ - $e_2$	miedo-ira	anticipación-sorpresa	anticipación-tristeza	miedo-miedo	sorpresa-alegría
$S_2$ -Octavas	4 - 4	5 - 5	5 - 6	6-6	6-6
$S_3$ -Act	0.053	0.03	0.038	0.057	0.048
$S_3$ -JS	0.0011	0.0004	0.0103	0.0236	0.0103
$S_3$ -Tempo	164	92	117	177	133
$S_3$ - $e_1$ - $e_2$	ira-alegría	alegría-miedo	ira-sorpresa	alegría-alegría	ira-miedo
$S_3$ -Octavas	4 - 6	6 - 6	4 - 5	6-6	4-6
$S_4$ -Act	0.043	0.033	0.026	0.058	0.039
$S_4$ -JS	0.0042	0.0007	0.0048	0.0236	-0.0066
$S_4$ -Tempo	133	94	79	180	110
$S_4$ - $e_1$ - $e_2$	alegría-anticipación	alegría-tristeza	tristeza-sorpresa	sorpresa-ira	tristeza - miedo
$S_4$ -Octavas	6-6	6-6	6-6	5-4	6-6

"*Peter Pan and Wendy*" difiere ampliamente (0.023), presentando una alta variabilidad en la actividad emocional. Por otro lado, el texto de "*Don Quixote of La Mancha*" (0.001) presenta el valor más bajo de variabilidad de actividad. Utilizamos todos estos valores para componer la pieza musical y generar la animación audiovisual que transmite con precisión las emociones extraídas del texto.

La Figura 5.3 muestra los grafos resultantes para las cuatro secciones de "*Peter Pan and Wendy*". Adicionalmente, en un pentagrama mostramos la parte inicial de la composición musical basada en las emociones extraídas. El gráfico que aparece en el extremo izquierdo (Figura 5.3a) muestra una mayor interacción entre Wendy y Peter en comparación con otros personajes, y es donde la alegría y la anticipación predominan como emociones principales. Sin embargo, en el gráfico superior derecho (Figura 5.3b) que representa la segunda sección de la obra, la relación más extensa involucra a Peter y John. En el inferior izquierdo del gráfico (Figura 5.3c), el protagonismo de los diálogos vuelve de nuevo a Wendy. Finalmente, la última sección es representada visualmente por la figura del extremo inferior derecho del gráfico (Figura 5.3d), la cual involucra el mayor número de personajes, como Liza y Cecc en la escena, y a su vez, predominan la alegría y la anticipación como emociones  $e_1$  y  $e_2$ , tal cuál sucedió en la primera sección.

En la Figura 5.4, procesamos todo el texto de "*Psycho*", sin dividirlo por secciones, para generar un resumen global del libro. Como se puede observar, el mayor número de diálogos involucra a Adam, seguido por Norma, Marty y Roy\_A, respectivamente. Este mapa global revela aspectos importantes, como personajes que nunca se relacionan entre sí (Barbara y Jesus, o Queen y Norma), o incluso personajes que dialogan con una sola persona a lo largo de toda la historia (Clara con Adam, o Queen con Paul). Sobre la composición musical resultante "*Psycho*", registra el menor tempo en la mayoría de sus secciones, provocando una velocidad lenta en comparación con el resto de las obras. La emoción predominante es el miedo, por lo que se eligieron progresiones de acordes que generan tensión y una escala menor, y se añadieron notas en la melodía fuera de la escala para añadir disonancia. Además, la elección de instrumentos como el *Glockenspiel* y *Orchestral Strings* acentúan el color de la emoción.

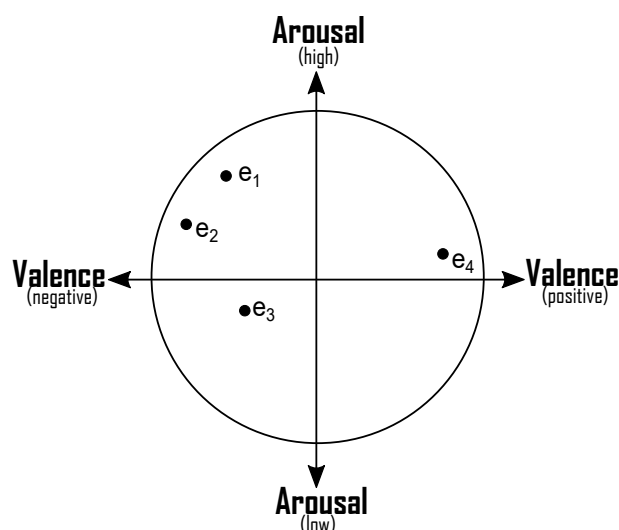
## 5.3. Estudio con usuarios

Realizamos dos estudios de usuarios para evaluar el desempeño de nuestra propuesta en términos de capacidad de síntesis y transmisión de emociones. Nuestra primera prueba valida la emoción que expresa la composición musical basada en los rasgos de *arousal* y *valence*. Nuestra segunda prueba tiene como objetivo medir la experiencia del usuario después de interactuar con nuestro prototipo.

### 5.3.1. Probando la composición musical

Utilizamos cuatro de nuestros resultados para confirmar si nuestra composición musical transmite la emoción extraída de las obras: “*Peter Pan and Wendy*” (*alegría*), “*Psycho*” (*miedo*), “*Don Quixote of La Mancha*” (*alegría*), y “*The three musketeers*” (*tristeza*). Cada una de estas secciones expresa la emoción específica descrita entre paréntesis. Utilizamos un esquema *arousal-valence* que nos ayuda a fijar la emoción del usuario en un mapa 2-dimensional,

ilustrado en la figura adjunta, donde eje-*Arousal* mide cuán tranquila (baja) o excitante (alta) es la información, y eje-*Valence* codifica los eventos emocionales como positivos o negativos. Luego, posicionamos cuatro puntos como posibles emociones para nuestro experimento, *i.e.*,  $e_1$  (■) es *miedo*,  $e_2$  (■) es *ira*,  $e_3$  (■) es *tristeza*, y  $e_4$  (■) es *alegría*. Reunimos a un grupo de 30 participantes, todos ellos estudiantes de pregrado, para interpretar las cuatro composiciones mencionadas. Luego, les pedimos que asignen cada composición al punto que consideren que encaja mejor. La Figura 5.5 resume los resultados producidos para cada caso.



Nótese que la mayoría de los participantes afirman percibir la emoción que pretendemos transmitir, es decir, un 70 % de *alegría*, un 83,3 % de *alegría* y un 73,3 % de *tristeza*, tal y como se muestra en las Figuras 5.5a, 5.5b y 5.5c respectivamente. Sin embargo, en 5.5d, aunque la mayoría de los candidatos (50 %) elige la emoción correcta (*miedo*), existe una porción (36.7 %) que elige la más similar a ella (*ira*). Representa una decisión difícil ya que ambas emociones se sitúan en el mismo cuadrante y tienen valores muy cercanos para los atributos de *arousal* y *valence*.

### 5.3.2. Preguntas sobre la experiencia del usuario usando nuestro prototipo

Logramos esta tarea reuniendo a siete participantes, distintos a los del estudio anterior, para leer una sección de “*Peter Pan and Wendy*” con una extensión de 35 páginas.



En seguida, les pedimos que respondan a las preguntas detalladas en la Tabla 5.2 sobre su opinión acerca de las capacidades y limitaciones de nuestro prototipo, donde el valor de uno ■ (1) calificaba como el más bajo y cinco ■ (5) como el más alto. La Figura 5.6 resume todas las respuestas de los participantes. Como se puede observar, Q1 y Q2 evalúan la composición musical y la animación basada en grafos por separado, y la mayoría de los participantes están de acuerdo en que ambos se desempeñan satisfactoriamente. Sin embargo, en Q3 les preguntamos sobre la percepción de nuestro prototipo, y los resultados mostraron un valor decreciente (tres personas clasificaron como medio el nivel de asimilación). Esta información es muy útil para nuestros propósitos de seguir investigando sobre este tema.

Tabla 5.2: Preguntas sobre la experiencia de usuario en nuestro segundo estudio de usuarios.

ID	Pregunta
Q1	<i>¿La composición musical en combinación con la visualización te permite generar un concepto sobre el contenido de la sección de la obra?</i>
Q2	<i>¿Considera que la animación basada en grafos ayuda a mejorar la comprensión de la interacción entre los personajes de la sección de la obra?</i>
Q3	<i>¿Considera que el prototipo resume la obra en una manera fácil de asimilar (5) o sobrecarga su percepción mediante el uso de animación y composición musical (1)?</i>

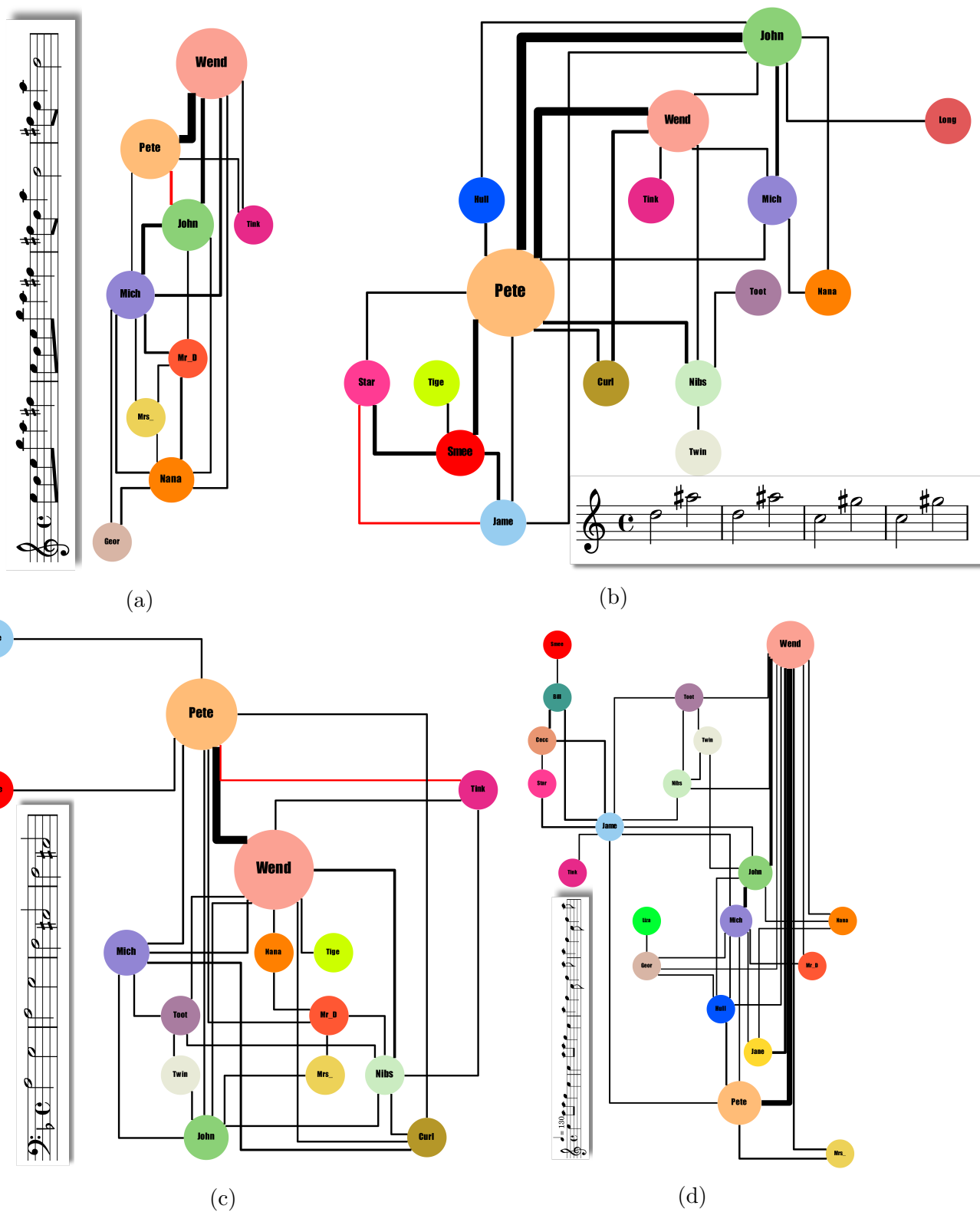


Figura 5.3: Mostramos los grafos resultantes de haber procesado las cuatro secciones de “Peter Pan and Wendy”.

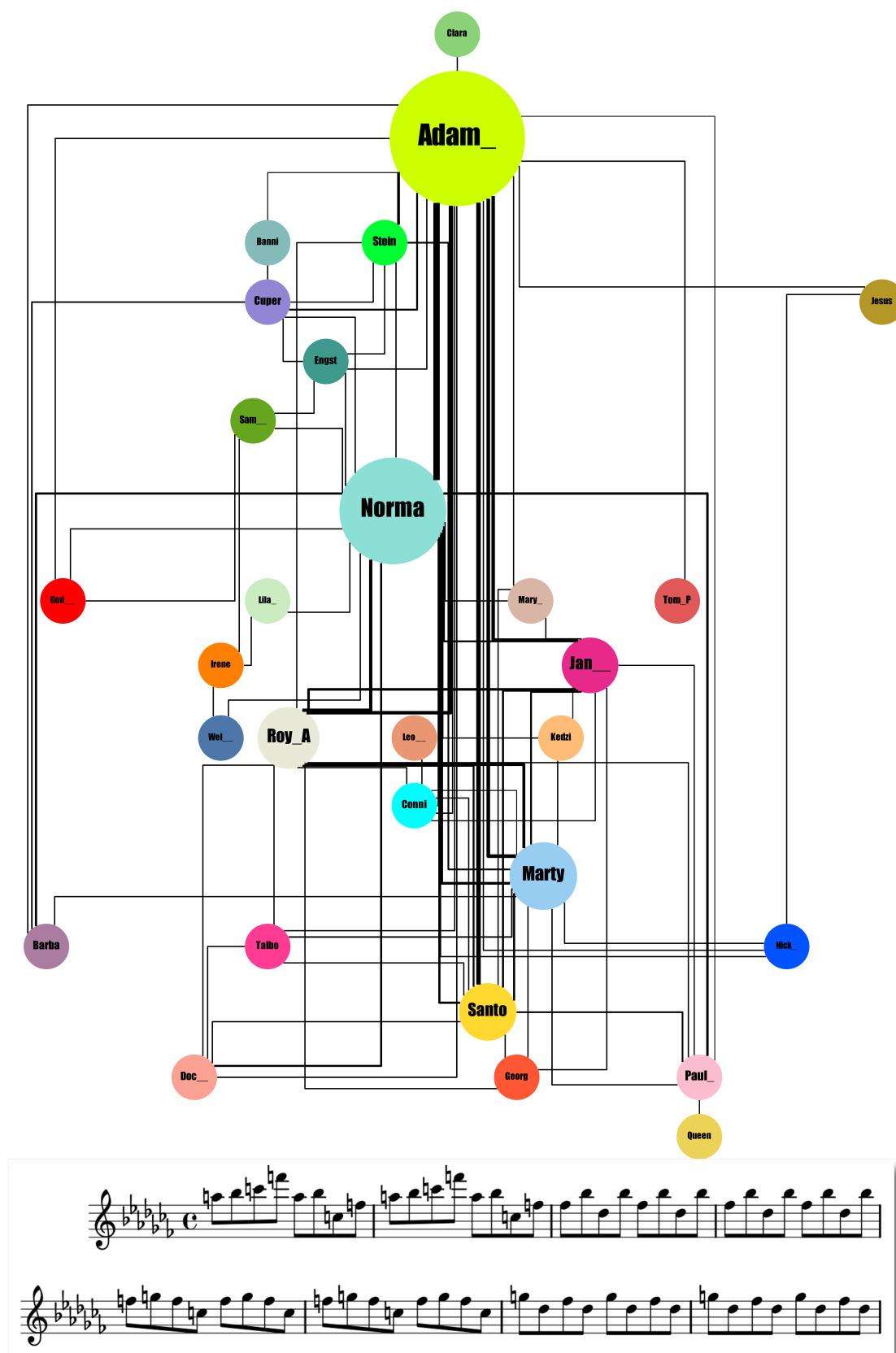


Figura 5.4: Representación visual del procesamiento completo de "Psycho" por Robert Bloch.

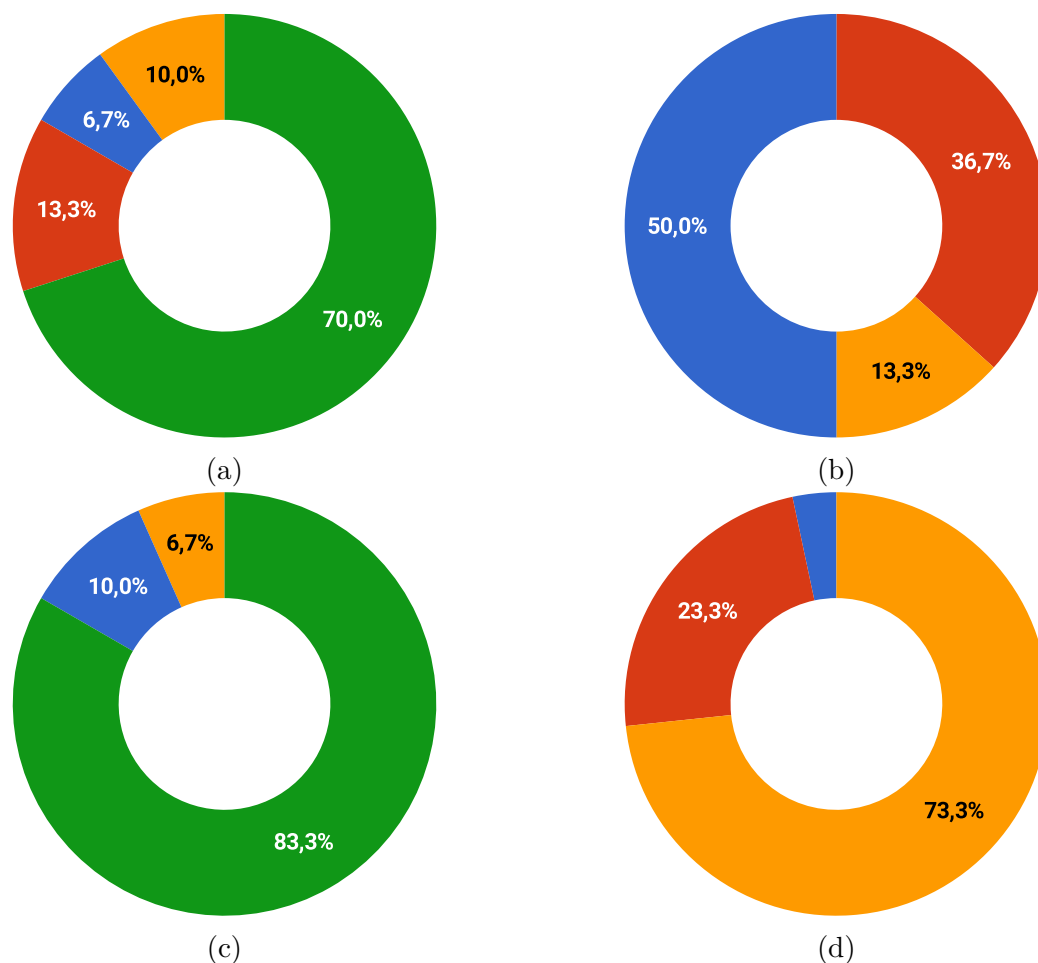


Figura 5.5: Resultados obtenidos por nuestro primer estudio de usuarios acerca de cuatro composiciones musicales diferentes: (a) “*Peter Pan and Wendy*” (alegría), (b) “*Psycho*” (miedo), (c) “*Don Quixote of La Mancha*” (alegría), y (d) “*The three musketeers*” (tristeza). Los colores representan la emoción elegida: ■ = miedo, ■ = ira, ■ = tristeza, y ■ = alegría por cada uno de los 30 participantes.

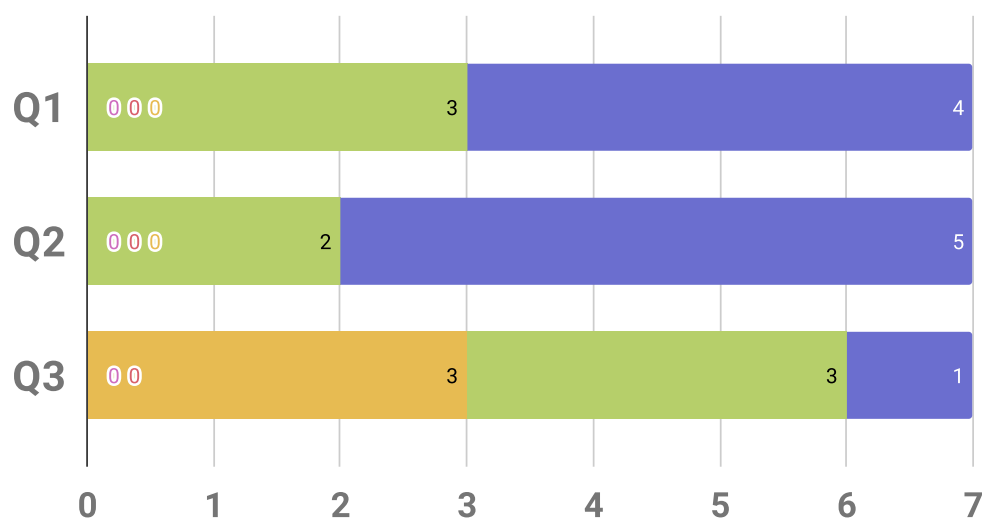


Figura 5.6: Resultados obtenidos por nuestro segundo estudio de usuarios en la primera sección de “*Peter Pan and Wendy*”. Los colores representan el valor elegido: ■ (1), ■ (2), ■ (3), ■ (4), y ■ (5) por cada uno de los 7 participantes.

## Capítulo 6

# Conclusiones y trabajos futuros

### 6.1. Conclusiones

En este trabajo se presenta una metodología para la generación de resúmenes audiovisuales a partir de obras literarias. Nuestro enfoque utiliza la extracción de emociones mediante el análisis léxico para generar composiciones musicales y animaciones basadas en grafos que transmiten dichas emociones. Nuestro prototipo tiene como objetivo, a través de los recursos multimedia, aumentar el compromiso de los usuarios para descubrir en detalle la historia que se esconde detrás de nuestro resumen.

Para alcanzar nuestros objetivos realizamos varios experimentos, el primero de ellos fue encontrar el método adecuado para realizar la extracción de emociones a partir de una obra literaria. Para abordar este objetivo, hicimos uso de dos recursos del Análisis de emociones en un texto: Texto etiquetado y Lexicones emocionales. Tal y como vimos en la sección de Resultados, aunque la intención era utilizar técnicas de aprendizaje profundo, la escasez de datos asociados a una emoción, ocasionó un mal desempeño en la red neuronal propuesta, provocando un desbalance al momento de la predicción de las emociones. Con respecto al segundo recurso empleado, decidimos usar NRC Lexicon, uno de los pocos lexicones emocionales que encontramos en la literatura. Con base a nuestras pruebas, notamos que al usar el *lexicon* en textos cortos como canciones y poemas, se generaban resultados pobres en comparación al emplear el mismo *lexicon* en textos con mayor longitud como obras literarias con múltiples capítulos (*Peter Pan and Wendy*: 17 capítulos). Por esta razón y por poseer una mayor variedad de emociones (alegría, tristeza, ira, miedo, anticipación y sorpresa) en comparación al primer recurso (alegría, tristeza, ira y miedo), para nuestro primer objetivo, decidimos utilizar NRC Lexicon para la extracción de emociones en una obra literaria.

Con las emociones extraídas, el siguiente objetivo fue generar un método que componga una pieza musical con múltiples instrumentos. Para ello implementamos un algoritmo basado en reglas. El algoritmo se encarga principalmente de componer una pieza musical de forma automática basándose en tres módulos primordiales: La armonía, el ritmo y la melodía. Combinando cada módulo es posible componer una pieza musical basada en una emoción específica. Se implementaron algoritmos para hallar la progresión de

acordes, el tempo, las octavas, adornos melódicos, selección de la escala, selección de instrumentos, entre otros. Así, con cada uno de estos componentes, fue posible ir moldeando la melodía en base a la emoción más predominante en cada sección de la obra literaria. El método dio como resultado cuatro piezas melódicas con múltiples instrumentos (una por sección) que reflejan las emociones de la obra literaria.

Luego, decidimos acompañar a la pieza melódica con un recurso visual para ofrecer una experiencia más placentera al usuario. De esta forma, fue necesario implementar un método para mapear los personajes y diálogos obtenidos de la obra literaria en un grafo animado. Para este paso, se tuvo que realizar una serie de subtarefas como: El preprocesamiento del texto, armar un *dataset*, realizar el conteo de emociones y el cálculo de densidades. Además se empleó la librería *Spacy* para detectar las entidades en la obra. Con *Spacy*, fue necesario corregir muchos errores manualmente, empezando por crear un conjunto de datos de los personajes más famosos de los cuentos de hadas como Pinocchio o Tinker Bell, ya que con esta librería no era posible detectarlos. También se implementó un algoritmo que nos permitía agrupar las correlaciones. Así cuando teníamos Miss Darling y Wendy Darling (el mismo personaje con diferente nomenclatura) el grafo solo mostraba una de sus correlaciones. Luego, habiendo culminado con este proceso, se obtuvo la lista de los personajes más relevantes de cada obra y la cantidad de diálogos en los que estos intervenían en un momento específico. Con esta información, se realizó la animación de un grafo, en donde los nodos representaban a los personajes de la obra y las aristas la relación entre cada personaje. El grosor de la arista y el nodo aumentaban a medida que el mismo personaje iba apareciendo más veces en la obra. Por último, se obtuvo el prototipo final, que fue la combinación del grafo animado y la pieza musical que se generó anteriormente.

Finalmente, decidimos realizar dos pruebas con usuarios para evaluar el desempeño de nuestra propuesta. Con el primer estudio, la intención fue probar qué tanto logra el modelo transmitir una determinada emoción al usuario. Por lo que les mostramos 4 piezas musicales correspondientes a distintas obras literarias e hicimos que ellos mismos ubicaran la emoción en el esquema *arousal-valence*. Con base a los resultados concluimos que los usuarios reconocieron correctamente la emoción en todos los casos. Para nuestro segundo estudio, decidimos evaluar la experiencia de nuestros usuarios al usar nuestro prototipo. Con estos resultados, nos dimos cuenta que la melodía musical en combinación con el grafo, brindaba a los usuarios un buen concepto general sobre la sección de la obra. Además, la mayoría coincidió con que el grafo los ayudaba a mejorar su comprensión con respecto a la interacción entre los personajes. Y el último *ítem* nos ayudó a concluir que la animación del grafo, más el texto de los diálogos y la melodía musical, sobrecargaba a los usuarios. Ambos, estudios nos permitieron recolectar información muy útil para seguir mejorando el modelo.

## 6.2. Trabajos futuros

Como parte de los trabajos futuros, proponemos algunas ideas para continuar con el desarrollo de nuestro prototipo y de esta forma, continuar brindando una experiencia placentera al usuario a través de nuestra metodología. Parte de estos nuevos objetivos, se enfocan en utilizar nuevas técnicas de aprendizaje de máquina para permitir que nuestro

modelo sea escalable y en realizar mejoras en la parte visual de nuestro diseño con el *feedback* obtenido en la prueba con usuarios.

1. Aplicar nuevas técnicas de aprendizaje de máquina para procesar una mayor cantidad de datos y alcanzar mayor escalabilidad.
2. Diseñar un aspecto visual que no sobrecargue al usuario.
3. Implementar herramientas interactivas para la exploración del usuario.
4. Modificar las características del prototipo para la generación de resúmenes audiovisuales que no se limiten a obras literarias.

Con estas nuevas ideas pensamos que el usuario podría fortalecer sus capacidades de análisis y así mejorar su legibilidad y comprensión lectora. Pero sobre todo, deseamos que su interés por leer obras literarias despierte luego de haber reproducido el resumen.

# Bibliografía

- E. C. O. Alm. Affect in text and speech, 2008.
- S. Aman and S. Szpakowicz. Identifying expressions of emotion in text. In *Proceedings of the 10th International Conference on Text, Speech and Dialogue*, TSD'07, pages 196–205, Berlin, Heidelberg, 2007. Springer-Verlag. ISBN 3-540-74627-7, 978-3-540-74627-0. URL <http://dl.acm.org/citation.cfm?id=1776334.1776363>.
- S. Aman and S. Szpakowicz. Using roget's thesaurus for fine-grained emotion recognition. In *In: Proc. Third International Joint Conf. on Natural Language Processing (IJCNLP)*, pages 296–302, 2008.
- M. C. Ardanuy and C. Sporleder. Structure-based clustering of novels. In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, pages 31–39, 2014.
- G. P. Basharin, A. N. Langville, and V. A. Naumov. The life and work of a. a. markov, 2004.
- H. Baxter and M. Baxter. *Cmo Leer Msica /The Right Way to Read Music*. 2007.
- S. Chaffar and D. Inkpen. Using a heterogeneous dataset for emotion analysis in text. In C. Butz and P. Lingras, editors, *Advances in Artificial Intelligence*, pages 62–67, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-21043-3.
- W. Chase. *How Music Really Works*. Roedy Black Publishing; Edition: Second Edition (2006), 2006.
- H. Chu, R. Urtasun, and S. Fidler. Song from pi: A musically plausible network for pop music generation. *arXiv preprint arXiv:1611.03477*, 2016.
- H. Davis and S. M. Mohammad. Generating music from literature. *CoRR*, abs/1403.2124, 2014.
- D. K. Elson, N. Dames, and K. R. McKeown. Extracting social networks from literary fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 138–147, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- F. Findeisen. *The Addiction Formula*. Albino Publishing, 2015.
- Gabrielsson and Lindström. The influence of musical structure on emotional expression. 2001.



- 
- J. C. García Torrecilla. Generación automática de música mediante tecnologías de aprendizaje profundo. 2018.
- K. Goel, R. Vohra, and J. K. Sahoo. Polyphonic music generation by modeling temporal dependencies. 2014.
- R. Hausser. *Foundations of Computational Linguistics: Human-Computer Communication in Natural Language*. Springer Publishing Company, Incorporated, 3rd edition, 2014. ISBN 3642414303, 9783642414305.
- E. Herrera. *Teoría musical y armonía moderna*. Antoni Bosch, 1984.
- L. A. Hiller Jr and L. M. Isaacson. Musical composition with a high speed digital computer. In *Audio Engineering Society Convention 9*. Audio Engineering Society, 1957.
- J. Hobbs. Musical key characteristics and emotions, 2018. URL <https://ledgernote.com/blog/interesting/musical-key-characteristics-emotions/>.
- B. L. Jacob. Algorithmic composition as a model of creativity. *Org. Sound*.
- R. L. B. M. E. F. James W. Pennebaker, Roger J. Booth. Linguistic inquiry and word. pages 1083–1086, 2001.
- Juslin and Västfjäll. Emotional responses to music: The need to consider underlying mechanisms. pages 559–575, 2008.
- M. Kennedy, J. Bourne, and J. Kennedy. *The Oxford dictionary of music*. Oxford University Press, 2006.
- F. Krebs, B. Lubascher, T. Moers, P. Schaap, and G. Spanakis. Social emotion mining techniques for facebook posts reaction prediction. *Proceedings of the 10th International Conference on Agents and Artificial Intelligence*, 2018. doi: 10.5220/0006656002110220. URL <http://dx.doi.org/10.5220/0006656002110220>.
- A. Latham. Diccionario enciclopédico de la música, 2009.
- E. Lindström and Västfjäll. Impact of melodic organization on perceived structure and emotional expression in music. pages 85–117, 2009.
- R. Masu, C. CORE, A. CONCI, F. MORREALE, A. DE ANGELI, et al. Robinflock: a polyphonic algorithmic composer for interactive scenarios with children. 14th Sound and Music Computing Conference, 2017.
- T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space, 2013.
- E. R. Miranda and J. Al Biles. *Evolutionary computer music*. Springer, 2007.
- S. Mohammad. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, LaTeCH '11, pages 105–114, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics. ISBN 9781937284046.

- S. M. Mohammad and F. Bravo-Marquez. Emotion intensities in tweets. In *Proceedings of the sixth joint conference on lexical and computational semantics (\*Sem)*, Vancouver, Canada, 2017.
- S. M. Mohammad and P. D. Turney. Crowdsourcing a word-emotion association lexicon. 29(3):436–465, 2013.
- F. Morreale, R. Masu, and A. De Angeli. Robin: An algorithmic composer for interactive scenarios. 08 2013.
- J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, 2014. URL <http://www.aclweb.org/anthology/D14-1162>.
- R. Plutchik. Emotions: A general psychoevolutionary theory. *Approaches to emotion*, 1984:197–219, 1984.
- J. D. F. Rodriguez and F. J. Vico. AI methods in algorithmic composition: A comprehensive survey. abs/1402.0585, 2014.
- R. Rowe. *Machine Musicianship*. Mit Press. MIT Press, 2001. ISBN 9780262681490.
- J. Salas. Generating music from literature using topic extraction and sentiment analysis. *IEEE Potentials*, pages 15–18, 2018.
- C. Schmidt-Jones and R. Jones. *Understanding Basic Music Theory*. Connexions, 2011.
- A. Seyeditabari, N. Tabari, and W. Zadrozny. Emotion detection in text: a review, 2018.
- I. Simon, D. Morris, and S. Basu. Mysong: Automatic accompaniment generation for vocal melodies. 2008.
- P. M. Sosa. Twitter sentiment analysis using combined lstm-cnn models. pages 1–9, 2017.
- C.-C. Stere and t. Trausan-Matu. Generation of musical accompaniment for a poem, using artificial intelligence techniques. *Computer Interaction*, 2017.
- M. Towsey, A. Brown, S. Wright, and J. Diederich. Towards melodic extension using genetic algorithms. *Educational Technology & Society*, pages 54–65, 2001.
- R. Valitutti. Wordnet-affect: an affective extension of wordnet. In *In Proceedings of the 4th International Conference on Language Resources and Evaluation*, pages 1083–1086, 2004.
- M. C. Waumans, T. Nicodème, and H. Bersini. Topology analysis of social networks extracted from literature. *PLOS ONE*, 10(6):1–30, 06 2015. doi: 10.1371/journal.pone.0126470. URL <https://doi.org/10.1371/journal.pone.0126470>.
- D. Williams, A. Kirke, E. R. Miranda, E. Roesch, I. Daly, and S. Nasuto. Investigating affect in algorithmic composition systems. *Psychology of Music*, 43(6):831–854, 2015.
- M. Yamada and Y. Murai. Story visualization of literary works. *Journal of Visualization*, 12(2):181–188, Jun 2009.

M. Yamada, Y. Murai, and I. Kumagai. Story visualization of novels with multi-theme keyword density analysis. *Journal of Visualization*, 16(3):247–257, Aug 2013. ISSN 1875-8975.

J. Zamacois. *Teoría de la música: dividida en cursos*. Labor, 1978.